Experimental evolution in *TP53* deficient human gastric organoids recapitulates tumorigenesis

Kasper Karlsson[1,2,3,#,†], Moritz Przybilla[3,%,†], Hang Xu[1,2,3], Eran Kotler[2,3], Kremena Karagyozova[3], Alexandra Sockell[2], Katherine Liu[4,3], Amanda Mah[5], Yuan-Hung Lo[5], Bingxin Lu[6], Kathleen E. Houlahan[1,2,3], Aziz Khan[3], Zhicheng Ma[3], Carlos J. Suarez[7], Chris P. Barnes[6], Calvin J. Kuo[5], Christina Curtis[1,2,3]*

[1]Department of Medicine, Division of Oncology, Stanford University School of Medicine, Stanford, CA, USA. [2]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [3]Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. [4]Department of Biology, Stanford University School of Medicine, Stanford, CA, USA. [5]Department of Medicine, Division of Hematology, Stanford University School of Medicine, Stanford, CA USA. [6]Department of Cell & Developmental Biology, University College London, London, UK. [7]Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA.
[#]Current address: Department of Oncology-Pathology, Karolinska Institute and SciLifeLabs, Stockholm, SE
[%]Current address: Wellcome Sanger Institute & University of Cambridge, Hinxton, Cambridgeshire, UK
[†]Contributed equally

*Correspondence: cncurtis@stanford.edu

Abstract

The earliest events during human tumor initiation are poorly characterized but may hold clues as to how to detect and prevent malignancy. Here we model occult pre-neoplasia by engineering *TP53* deficiency, a common and early event in gastric cancer, into human gastric organoids. By performing experimental evolution in multiple clonally derived cultures over two years we define causal relationships between this common initiating genetic lesion and resulting phenotypes**.** *TP53* loss elicited progressive aneuploidy, including copy number alterations and complex structural variants that are common in gastric cancers and which follow preferred temporal orders. Longitudinal single cell sequencing of *TP53* deficient gastric organoids similarly indicates progression towards malignant transcriptional programs. Moreover, lineage tracing with expressed cellular barcodes demonstrates reproducible dynamics whereby initially rare subclones with shared transcriptional programs repeatedly attain clonal dominance. This powerful platform for experimental evolution exposes stringent selection, clonal interference and a striking degree of phenotypic convergence in pre-malignant epithelial organoids, implying that the earliest stages of tumorigenesis may be predictable while illuminating evolutionary constraints and barriers to malignant transformation.

Introduction

In rapidly adapting asexual populations, including microbes, viruses and cells within tumors, multiple mutant lineages often compete for dominance within the population (*1*). These complex dynamics determine the outcomes of evolutionary adaptation, but are difficult to directly observe *in vivo* (*2*). Experimental evolution *in vitro* can provide a window into this otherwise unobservable process, enabling characterization of adaptation, the identification of mutant clones and their fitness benefits (*3*, *4*). Such controlled experimental systems have yielded fundamental insights into microbial and viral evolution in unvarying or consistently varying conditions. The same forces of mutation and selection fuel clonal expansions in somatic cells during aging, contributing to malignancy, but the dynamics are poorly understood (*5*–*7*).

Cancers often arise from a mutated cell that undergoes premalignant clonal expansion, all the while accruing additional mutations. These somatic mutations accumulate and spread in phenotypically normal tissues prior to apparent morphological changes, with aneuploidy and driver gene mutations preceding cancer diagnosis by years (*5*, *6*, *8*–*10*). Identifying the causes of malignant transformation requires characterization of abnormal molecular phenotypes that precede this event in a tissue-specific manner. However, repeat sampling of healthy human tissue or pre-neoplasia is not practical. Critically, neither the order of somatic alterations nor the patterns of clonal expansion that precede malignant transformation can be inferred from genomic sequencing of established tumors (*5*, *10*). New approaches are therefore needed to characterize occult pre-neoplastic clonal dynamics to inform strategies for earlier detection and intervention.

Identification of the drivers and dynamics of pre-malignancy is especially critical for cancers that have long lead times but lack routine screening, contributing to late diagnoses and poor clinical outcomes. As the fourth-leading cause of cancer mortality worldwide, gastric cancer (GC) epitomizes this challenge and represents a major public health burden due to late presentation, poor prognosis and limited treatment options (*11*, *12*). Accordingly, there is a critical need to identify the molecular determinants of GC and its precursors. Recent efforts have resulted in forward-genetic and tumor-derived murine and human GC organoids (*13*–*16*), demonstrating their utility as pre-clinical models. Although forward-genetic models can capture aspects of pre-neoplasia, combinatorial hits are often engineered to accelerate tumorigenic phenotypes, bypassing nascent progression (*13*, *14*).

We take a distinct approach to model tumorigenesis from the “bottom up”, harnessing the power of experimental evolution, human organoids and CRISPR/Cas9-mediated genome editing to define causal relationships between initiating genetic lesions and the resultant genotypes and phenotypes by tracing somatic alterations and clonal expansions over multiple generations *in vitro*. Using human gastric organoids as a *tabula rasa*, we study pre-neoplasia by engineering *TP53* inactivation in replicate clonally derived cultures from non-malignant tissue from three donors and performing experimental evolution for over two years under defined conditions. Organoids recapitulate cellular attributes of *in vivo* models, including three-dimensional tissue structure, multi-lineage differentiation and disease pathology, while retaining the tractability of *in vitro* models, rendering them ideal for this task (*17*). We model *TP53* deficiency since this potent tumor suppressor is altered in the majority (>70%) of chromosomal instable (CIN) GCs, the largest molecular subgroup of disease, comprising 50% of all GCs. Inactivation of *TP53* is a nearly ubiquitous early event preceding numerical and structural chromosomal abnormalities (aneuploidy) in CIN GC (*14*). While most solid cancers are aneuploid due to a CIN phenotype with p53 loss permissive for CIN (*18*), whether *TP53* loss can elicit aneuploidy remains controversial and is likely tissue type dependent (*19*–*22*). For example, *TP53* deficiency or the introduction of *TP53* hotspot missense mutations was sufficient for aneuploidy in non-transformed MCF10a mammary cells, whereas in colon organoids, dual inactivation of the *APC* gatekeeper tumor suppressor and *TP53* was needed, emphasizing tissue specificities and context dependencies (*23*–*25*). More generally, *TP53* alterations are associated with aneuploidy and poor prognosis across human cancers, irrespective of mutation type (truncating versus missense), but the extent

to which specific aneuploidies and copy number alterations (CNAs) are selectively advantageous and their functional impact on transcriptional programs and malignant transformation has been difficult to study (*18*, *21*).

Here we investigate the molecular and phenotypic consequences of this common initiating event through longitudinal genome sequencing, transcriptional profiling and cellular barcoding of independent *TP53* deficient (null) gastric organoid cultures derived from three donors. Given the varied functions of the p53 network (*26*, *27*), we model *TP53* deficiency rather than specific loss or gain of function point mutations, facilitating interpretation of resultant phenotypes. We demonstrate that multi-year experimental evolution in *TP53*$^{-/-}$ gastric organoids recapitulates genomic and transcriptomic hallmarks of gastric malignancy, including the multi-hit, temporal acquisition of CNAs, enabling unbiased discovery of the molecular determinants of adaptation to *TP53* deficiency and candidate drivers. The resultant data define genotype to phenotype maps of gastric pre-neoplasia, revealing extensive clonal interference and rapid adaptation, underpinned by temporal genomic contingencies and phenotypic convergence. Our findings highlight the power of experimental evolution in human organoids to investigate pre-neoplastic processes and the repeatability of somatic evolution with implications for anticipating and thwarting malignancy.

TP53 loss induces gastric cancer associated CNAs in a defined temporal order

To model the earliest stages of tumor initiation in CIN GC, we established human gastric organoids from non-malignant corpus (body) tissue from three donors undergoing sleeve gastrectomy and performed CRISPR/Cas9-editing to introduce bi-allelic *TP53* frameshift-inducing indels, thereby yielding an inactive gene product (**Fig. 1A, Fig. S1-2, Methods**). For each donor (D1-3), 3 independent clonally-derived cultures (C1-3) were established, yielding 9 cultures for long-term propagation, 5 of which were further split into 3 replicates each (R1-3) for barcoding studies (n=24 cultures, **Table S1, Methods**). A second 'hit' in the tumor suppressor *APC,* a negative regulator of the Wnt pathway, altered in 20% of CIN GC (**Fig. S2A,B**) was engineered concurrently in cultures 2 and 3 from donor 3 (referred to as D3C2 and D3C3, **Fig. S2C,D**) to investigate whether double knockout (KO) promotes accelerated evolution or aneuploidy. Single CRISPR/Cas9-edited organoids were picked following selection with Nutlin-3 (for *TP53* loss) and removal of *WNT* and *RSPO* (for *APC* loss) and expanded (*25*). Clonal status of CRISPR edit sites were verified via Sanger sequencing and confirmed through WGS at multiple time points (**Fig. S2C,D, S3**). Throughout, we refer to time as days after *TP53* or *TP53*/*APC* KO unless otherwise stated.

We asked whether *TP53* deficiency elicits aneuploidy in gastric organoids, as evidenced by numeric and/or structural chromosomal abnormalities (genome instability). To investigate the landscape of CNAs, we performed shallow whole genome sequencing (sWGS; median 0.2x coverage, **Table S2**) across the 9 cultures at up to 7 timepoints each (n=125 sWGS assays). *TP53* deficient organoids progressively acquired CNAs, where chromosome-arm level losses accrued first, followed by copy number gains (**Fig. 1B**). In contrast, wild-type (WT) gastric organoids remained genomically stable during long-term culture (13-26 passages, **Fig. S4A**). Distinct donors and *TP53* KO cultures from the same donor exhibited variable patterns of CNAs, suggesting that genetic background does not wholly constrain the spectrum of subsequent alterations (**Fig. 1C, S4A**). Despite this variability, CNAs prevalent in the TCGA GC cohort (*28*) were recurrently altered in *TP53*$^{-/-}$ cultures, including loss of chromosome (chr) 3p (44% of organoid cultures, 32% in GC ie 44%, 32%), loss of chr 9p (77%, 44%), loss of chr 18q (67%, 31%) and gain of chr 20q (55%, 79%) (**Fig. S4B-C**). Additionally, arm-level CNAs present in two or more *TP53* or *TP53/APC* KO cultures were significantly more frequent than alterations present in one or fewer cultures in both gastric and esophageal cancers but not in other tumor types ($p < 0.05$, two-sided Wilcoxon rank sum test, **Fig. S4C**).

Across all cultures the fraction genome altered (FGA), a measure of aneuploidy, increased over time although the rates varied and tended to plateau after 600 days (**Fig. 1D**, **Methods**). For

example, D1C1, which accrued early arm-level alterations exhibited greater than 20% of genome altered by day 260, whereas the median FGA across all cultures at a similar midpoint was ~5%. Of note, the *TP53* KO and *TP53*/*APC* double KO cultures exhibited comparable levels of FGA at the final time point (average 11.3% and 10.7%, respectively), consistent with expectation that APC loss does not fuel aneuploidy in gastric cells and distinct from the requirement for dual *TP53/APC* KO for aneuploidy in colon organoids (*25*). As expected, these values are lower than that of established CIN, *TP53* mutant GCs in TCGA (median FGA 34.5% per cBioPortal). In several cultures, the FGA decreased over an interval, owing to clonal extinction. For instance, D3C3 had higher FGA at day 190 (7.8%) than day 442 (4.9%) due to the loss of a clone harboring chr13 deletion between days 190 and 260, as evident from the sWGS data (**Fig. 1D, S4A**). Similarly, D2C2 lost a subclone with chr9q amplification between days 428 and 609, resulting in a more modest decrease in FGA (16.4% to 14.5%).

We next investigated the temporal onset of arm-level and focal CNAs in *TP53* deficient gastric organoids, revealing constraints in the timing of these events (**Fig. 1E, Table S3**). For example, loss of chr9p and chr3p, repeatedly occurred (across donors and cultures) within 200 days after *TP53* KO but seldom later, suggesting a period during which these alterations were particularly selectively advantageous. The chr9p deletion spans the tumor suppressor *CDKN2A* which is commonly altered in the CIN subgroup of gastric (~41%, **Fig. S2A**) and esophageal (~74%) adenocarcinomas, and commonly co-occurs with *TP53* alterations (*14*, *28*). Loss of *CDKN2A* is also one of the earliest alterations associated with progression of Barrett's esophagus (BE) to dysplasia and esophageal adenocarcinoma (*29*). Moreover, in murine gastric organoids, co-deletion of *TP53* and *CDKN2A* promoted premalignant progression and induced replication stress, a known source of DNA breaks (*14*). Deeper sequencing of gastric organoids confirmed bi-allelic loss of *CDKN2A* via focal deletion (D3C1, D3C3) or truncating mutations in p16 (INK4A) (D1C3) on the backdrop of heterozygous loss (**Fig. S5A, S6B,D, Table S3**). Deletion of the fragile histidine triad protein (FHIT/FRA3B) encoded on chr3p also commonly occurred in *TP53* deficient gastric organoids by 190 days (median) as a result of whole arm (exemplified in D1C1, D2C2) and/or focal (exemplified in D3C1, D2C1) losses (**Fig. 1E, 2D, S6A,C, S9**, **Table S3**). Similar to *TP53, FHIT* functions as a genome caretaker that is often lost early during tumor progression and results in deoxythymidine triphosphate depletion contributing to replication stress and DNA breaks (*30*). Of note, 12% of CIN GC harbor *FHIT* alterations (**Fig. S2A**). The recurrent and apparent stringent selection for loss of these two tumor suppressors early during *in vitro* evolution of *TP53* deficient gastric organoids and in gastroesophageal tumors implies a role in tumor initiation, although alone they are insufficient for malignant transformation (*31*, *32*). Numerous additional GC-associated CNAs accrued during *in vitro* culture, including chr18q and gain of chr20q which consistently occurred relatively late (~600 days after *TP53* loss). Such preferentially late alterations may reflect changes in selective pressures concomittant with increased fitness or new evolutionary paths enabled by earlier alterations (*4*). These data demonstrate that *TP53* loss is permissive for aneuploidy in gastric cells, resulting in the acquisition of multiple GC-associated CNAs in a defined temporal order, reflecting evolutionary contingencies in the context of this common instigating event.

***TP53* deficiency elicits complex structural variation and clonal interference**

To evaluate genomic evolution following *TP53* loss in greater depth, we performed WGS (mean coverage 26x, **Table S4**) of five *TP53* deficient gastric organoid cultures sampled at *Mid* (296 days) and *Late* (718-756 days) time points, while two of these cultures (D2C2, D3C1) were sequenced at four additional time points (**Fig. 2A, S5A**). The sequencing data confirmed bi-allelic inactivation of *TP53* and *APC* at the expected CRISPR/Cas9 edit sites (**Fig. S3**). Similar to the findings based on sWGS, these data reveal an increase in the weighted genome instability index (wGII), the fraction of genome with loss of heterozygosity (LOH) as well as increased focal deletions and amplifications during prolonged culture (**Fig. 2A**, **S5B, Table S5**). These changes

were dominated by deletions while overall genomic content remained diploid as inferred from the genomic data and verified via FACS (**Fig. S5C, Table S6**). In addition, the WGS data enabled evaluation of the landscape of single nucleotide variants (SNVs) and structural variants (SVs), both of which increased with time (**Fig. 2A, S5B, Table S5**). Of note, the *TP53/AP*C double KO cultures (D3C2, D3C3) exhibited increased SNV burden relative to *TP53* KO at late timepoints, but in both cases, SNVs primarily occurred in intergenic and intronic regions, consistent with the lack of recurrent driver mutations in CIN GC (*28*).

Several regions of densely clustered mutations were noted, including those localized to the *FHIT* fragile site (**Fig. S7A, S7B**), which was also frequently deleted (**Fig. 1**). In fact, all cultures investigated by WGS exhibited *FHIT* hypermutation, although the location varied (**Fig. S7A,B**, **Table S3**). While all WT cultures had stable diploid genomes (**Fig. S4A**), for two donors (D1, D3) simple focal *FHIT* deletions were observed at late passages in the WT culture, likely due to clonal expansion of an initially rare event and suggestive of somatic mosaicism (**Fig. S6A,C**). We also investigated whether specific mutational processes were operative during *TP53*-driven genome evolution by examining known single base substitution (SBS) mutational signatures (*33*). The most prevalent mutational signatures were SBS 1, 5 and 40, which are ubiquitous and implicated in aging and cancer. However, by the *Late* time point, D3C2 developed SBS 17a and 17b (**Fig. S5B, Table S5**), which are prevalent in gastroesophageal adenocarcinomas and in dysplastic and non-dysplastic BE patients who subsequently progressed to invasive disease (*32*).

While all classes of alterations accumulated over time, the landscape of SVs was particularly notable, since at early timepoints deletions dominated, whereas duplications, translocations and inversions accrued by later time points (**Fig. 2A**). This is further exemplified in D3C1, looking across three timepoints where multiple inter-chromosomal rearrangements evolved with time (**Fig. 2B**) Such complex SVs have seldom been reported in normal tissues. Intriguingly, the *FHIT* locus harbored complex SVs, including rigma, which are deletion chasms occurring at fragile sites (**Fig. 2C**, **Fig. S8A, S9C**, **Methods**) recently reported in GC (**Fig. S9D**) and BE (*34*). Longitudinal WGS allowed us to trace the genesis of this rearrangement in D3C1, starting from a small deletion at the *FHIT* locus evident at day 115, with progressive breaks and complexity arising through multiple cell divisions, culminating in rigma by day 264. The subclone harboring this rearrangement was lost (**Fig. 2D-E,** yellow subclone) and a separate subclone (blue) with a distinct *FHIT* rigma emerged and persisted to the last timepoint, suggesting convergent evolution for this alteration. Notably, these rearrangements with multiple junctions evolve over several generations rather than as a single event, as previously proposed (*35*), and likely perpetuated by replication stress at this locus. Complex SVs evolved in other cultures (**Fig. S8A, S9C**), including a chr3 and chr9 translocation, which was verified by Sanger sequencing (**Fig. S8B-E**). Thus, evolved *TP53* deficient gastric organoids recapitulate multiple genomic hallmarks of gastrointestinal malignancy.

We further analyzed the longitudinal WGS data to investigate clonal competition and extinction in *TP53* deficient gastric organoid cultures by determining subclonal populations from CNA profiles (defined here via bulk WGS) across five timepoints for D3C1 and D2C2. By day 115, D3C1 had already acquired numerous deletions (9p, *FHIT*) and several SVs, including a translocation between chr11 and chr14, that persisted until the final time point. Over 600 days in culture multiple CNA-defined subclones increased in frequency before going extinct (**Fig. 2D,E**, **Table S7**). For example, a chr4$^-$, 9q$^+$ subclone arose early but went extinct by day 264, outcompeted by a chr19p$^-$ subclone that later acquired multiple alterations including chr8p$^-$, 9q.2$^+$, 16p$^-$ and remained dominant until day 404. This subclone was ultimately outcompeted by a subclone with loss of chr18q, which acquired gain of chr20q, a recurrent event in multiple cultures at late time points, followed by additional alterations. Thus, some clones fix and achieve dominance while others reach substantial frequencies before going extinct, likely due to clonal interference, which is common in asexual populations with a large supply of mutations and strong selection (*36*). Additionally, distinct CNA subclones co-exist at similar frequencies for extended

durations (~140 days or 10 passages) suggesting comparable fitness (*e.g.* chr8p$^-$, 9q.2$^+$,16p$^-$ and 18$^-$ subclones). Intermittent periods of clonal competition and stasis (~70 days) were similarly observed in other cultures (as shown for D2C2 in **Fig. S9A-B**). These dense longitudinal WGS data reveal the rise and fall of subclones, exposing stringent selection and pervasive clonal interference in premalignant epithelial populations. Motivated by these observations and recognizing that low frequency subclones can be obscured in bulk data, we investigated these patterns at single cell resolution via prospective lineage tracing, as detailed below.

Longitudinal single cell profiling reveals transcriptional deregulation following *TP53* loss

We next sought to investigate phenotypic and transcriptional changes during *in vitro* evolution of *TP53* deficient gastric organoids. To this end, we evaluated the growth dynamics of individual cultures and performed single-cell RNA sequencing (scRNA-seq) at *Early* (~100 days after *TP53* KO), *Mid* (323-413 days) and *Late* (588-862 days) time points (**Fig. 3A, S5A, Table S1**; **Methods**). We first investigated whether cell proliferation increased with time (passage number) by fitting a LOESS (locally-estimated scatterplot smoothing) regression model to cell numbers at each passage, where the growth derivative and fold change are a surrogate for fitness. Higher growth derivatives were observed at *Late* or *Mid* versus *Early* time points (**Fig. 3B, S10A**); similar results were obtained when raw cell numbers were considered (two-way repeated measures ANOVA, p-value = 0.003, **Fig. S10B**).

Across the 12 cultures (9 *TP53*-deficient, 3 WT), a total of 31,606 cells were retained for analysis following quality control assessment (**Fig. 3C,D,E**, **Table S8; Methods**). Since organoids recapitulate cellular populations from the tissue of origin (*37*), we first assessed the expression landscape of WT gastric organoids in relation to normal gastric tissue. Expression of normal gastric cell type markers were evident in WT organoid cultures from each of the three donors, including pit mucosal cell markers (PMCs: *MUC5AC*, *TFF1*, *TFF2*, *GKN2*) in D1, enterocyte markers (*FABP1*, *FABP2*, *ANPEP*, *PHGR1*, *KRT20*) in D2 and gland mucosal cell (GMCs) markers (*MUC6*, *PGC*, *TFF2*, *LYZ*) in D3, despite low cell numbers (322, 1689 and 44 cells for D1-D3 respectively) (**Fig. 3F-H, S11, Table S9**) (*22–25*). In contrast to WT cultures, the expression of gastric cell markers was heterogeneous in the *Early, Mid* and *Late* cultures *TP53* deficient cultures (**Fig. 3D,F,G; S11A,C,E; Table S10**), suggesting early and persistent transcriptional deregulation.

For D1 and D3, the mucosal-like phenotype in WT cultures, defined by the expression of mucin and TFF genes was lost following *TP53* KO. Additionally, in D1, intestinal goblet cell-specific markers, including *TFF3*, *WFDC2* and *MUC5B*, were upregulated at the *Late* time point, potentially indicative of intestinal metaplasia, a non-obligate precursor to GC (*38*). GC associated genes, including the claudins (*CLDN3*, *CLDN4*, *CLDN7)* and carcinoembryonic antigens (CEA) family (*CEACAM5*, *CEACAM6)* increased in expression over time in D1 and D3, while the inverse was observed in cultures from D2. These divergent expression patterns are due to the predominance of enterocytes in the initial D2 WT culture, likely attributable to inflammation in the gastrectomy biopsy (*39*) given the non-malignant origin and normal karyotype (**Fig. S4**). As expected, the CRISPR/Cas9 induced frameshift mutations in *TP53* and *APC* resulted in reduced expression of these genes, although residual expression can occur (*40*). Cultures harboring alterations in *CDKN2A* (D1C3, D3C1, D3C3) or *FHIT* (all except D1C2 and D1C3) exhibited low expression of these genes, which were also lowly expressed in WT cultures. Interestingly, some cultures exhibited increased *CDKN2A* expression (D1C1, D1C2 mid/late, C2 mid/late' D2C2 early/late; D3C2 early/late), presumably as a form of compensation to *TP53* inactivation, as similarly observed in murine models and some GCs (*14*).

We investigated the extent to which *TP53* deficient cultures share transcriptional features by intersecting significantly DEGs (adj. pval < 0.05, Wilcoxon Rank Sum test, Bonferroni correction) from *Early* to *Late* time point across the six cultures with scRNA-seq data available from an early timepoint. These analyses identified 13 consistently up-regulated and 40 down-

regulated genes, respectively (**Fig. 3I**, **Table S11**). Significantly upregulated genes include *CLDN4, TM4SF1*, *ZFAS1* each of which have been implicated GC (*28*, *41*). Significantly downregulated genes across these six cultures, included *LYZ* and *TFF2*, which modulate mucin production underscoring the general decrease in mucosal phenotypes in *TP53* deficient gastric organoids and *SPTBN1*, a cytoskeletal protein involved in TGF-β signaling (*42*).

Lastly, we assessed pathway-level changes by performing gene set enrichment analysis (GSEA) of differentially expressed genes (DEGs) between *Late* versus *Early* as well as *Mid* versus *Early* time points (**Fig. 3J, Table S12**). Several pathways were enriched across multiple cultures and donors, including upregulation of TNF-α signaling via NF-κβ as seen in comparisons of GC vs normal tissue (4/6 cultures) (*43*), as well as increased apoptosis (5/6 cultures) and hypoxia (5/6). Previous studies have demonstrated crosstalk between NF-κβ and p53 (*44*) consistent with the increased expression of NF-κβ following *TP53* loss observed here. Additionally, p53 negatively regulates protein levels and activity of *HIF1*α and a recent study demonstrated that mutant p53 induces a hypoxia transcriptional program during the initiation of mutant p53 gastric cancer xenografts (*45*). Downregulated pathways included MYC, E2F targets and G2M checkpoints, although this profile was more variable. Thus, despite the heterogenous trajectories observed at the single gene level, pathways implicated in malignancy were commonly altered across cultures and donors.

Unsupervised assessment reveals evolution towards malignant transcriptional states

To identify potential pathologic features within the evolved *TP53* deficient gastric organoids, we next compared them with normal and malignant gastric tissues. To this end, we performed an unsupervised analysis by projecting the organoid longitudinal scRNA-seq data onto a reference single cell atlas derived from GC patients using a computational framework based on latent semantic indexing (LSI, **Methods**) (*46*). The use of a reference enables comparisons across experimental conditions without requiring the definition of discrete cell types in the organoid dataset (*47*). We utilized a single cell GC dataset from Sathe et al. (*48*) comprised of 47,147 cells from eight tumors and adjacent matched normal as the reference atlas. Given the lack of stromal/immune cells in the organoid models, we restricted the atlas to cells expressing epithelial cell markers (*EPCAM* and *KRT18*) and confirmed the expression of *CD4*, *VIM*, *ACTA2* or *PTPRC* in excluded non-epithelial cells. This procedure yielded 6,001 epithelial cells (1,354 normal; 4,647 tumor; mean detected genes per cell = 1,964) which were assigned to distinct cell type clusters using literature-derived marker genes (*43*, *48*, *49*) (**Methods**). As in Sathe et al., two tumor cell clusters were apparent, which we termed mucosal-like malignant and non-mucosal-like malignant cells (*48*) because the latter included malignant markers *(KRT17*, *KRT7*, *LY6D*) but lacked mucosal markers (*MUC5AC*, *TFF2*, *TFF1*) (**Fig. 4A**). In addition, we assigned clusters to PMCs, GMCs, chief cells, parietal cells, enterocytes, enteroendocrine cells, goblet cells and proliferative cells (**Fig. 4B**).

We next projected batch corrected scRNA-seq data from *Early*, *Mid* and *Late* time points individually into the reference LSI subspace to identify the gastric cell types that are most similar (**Fig. 4C, S12, Methods**). Shifts in cell populations between time points were evident for all cultures, although for D1C2 and D1C3 this mainly occurred between two mucinous cell states. Some of these changes have been proposed to accompany the transition from normal tissue to gastritis which can lead to intestinal metaplasia and ultimately malignancy (illustrated schematically in **Fig. 4D**). To quantify changes in cell type frequencies for the individual cultures over time, we assessed the identity of the 25 closest nearest neighbors (NNs) in the reference population (**Fig. 4E**). These comparisons indicate an increase in mucosal-like malignant cells in 3/7 cultures at the *Late* time point, with 68.7%, 80.1% and 37.3% of NNs being mucosal-like malignant cells for D3C2, D3C3 and D1C1, respectively. In contrast, cultures from donor 2 (D2C2, D2C3) showed a strong decrease in mucosal-like malignant cells and an increase in non-mucosal-like malignant cells (D2C2 – 45.6%, D2C3 – 64.4%, NNs) from WT to the *Late* timepoint

(**Fig. 4E,** D2C1 lacked data at the late time point but exhibits similar patterns), explaining transcriptional differences relative to Donors 1 and 3 (**Fig. 3**). Notably, ~30% of cells in D2WT were located near enterocytes, potentially contributing to gastritis-like features, as noted above, and underlining the transcriptional similarity between enterocytes and malignant cells (*49*). In line with the supervised analysis above based on known gastric markers, WT cultures from D1 and D3 exhibited predominantly mucosal phenotypes. All cultures showed a decrease in PMCs at *Late* time points compared to the WT, *Early* and *Mid* time points. However, two D1 cultures were primarily located in non-malignant compartments, near GMCs (D1C2 – 73.3%, D1C3 – 41.8%, *Late*) a cell type prevalent in WT.

The decrease in mucosal genes and increase in GC-associated expression profiles suggests that the evolved *TP53* and *TP53*/*APC* deficient organoids are *en route* towards intestinal metaplasia and malignancy, albeit at different rates and through potentially different paths, corroborating the supervised analyses based on specific marker genes, detailed above. Intriguingly, while these cultures exhibit genomic and transcriptional hallmarks of CIN GC, including aneuploidy and intestinal markers, they do not exhibit evidence of histologic transformation (**Fig. S13**). This is consistent with the observation that somatic alterations accumulate and spread in phenotypically normal gastrointestinal mucosal epithelia years before appearance of identifiable morphological changes (*9*).

Deterministic outgrowth of rare subclones revealed by lineage tracing

We next sought to leverage our oncogene-edited organoid models to characterize subclonal dynamics during pre-neoplasia at higher resolution by exploiting high-complexity cellular barcodes. In contrast to retrospective lineage tracing via WGS, prospective lineage tracing via cellular barcoding enables the simultaneous detection of many low frequency subclones within the population, as well as estimation of their appearance times. Building on this principle but with the goal of jointly recovering lineage and transcriptional states, we developed expressed cellular barcodes (ECB) by engineering a strong mammalian promoter (EF1alpha) upstream of a 30bp semi random nucleotide sequence with a polyA-tail downstream (**Methods**, **Fig. S14A,B**). This high complexity GFP-tagged ECB library stably integrates into genomes via lentiviral transduction, such that each cell is uniquely labeled, facilitating lineage tracing of the progeny by a PCR reaction followed by barcode sequencing and linked scRNA-sequencing (via capture of the polyA-tail).

In total six *TP53* or *TP53*/*APC* deficient cultures (D1C1, D1C3, D2C1, D2C2, D2C3 and D3C2) were transduced with ECB lentivirus between days 101-115 and evolved in parallel to the non-barcoded cultures for more than one year (**Methods**). Each ECB parental line was split into three replicates, enabling evaluation of the reproducibility of clonal dynamics, where outgrowth of the same subclone is assumed to reflect an *intrinsic* fitness advantage, while divergent subclone dominance would suggest *acquired* fitness differences (**Fig. 5A**). As for the nonbarcoded samples, replicate cultures were Sanger sequenced for the *TP53* and *APC* deletion sites at multiple time points to verify clonality (**Fig. S14C**). For D1C3 a clone harboring a different *TP53* KO site grew out and hence was excluded (**Fig. S14D**).

Longitudinal sWGS of these long-term ECB cultures demonstrated a striking degree of reproducibility at the genomic level with recurrent CNAs shared across replicate cultures (**Fig. 5B, S15-16**). For example, in D2C2, new CNAs emerged around day 258 (loss of chr4q and chr13; gain of chr20q) across all three replicates. In D2C1, replicate 2 (D2C1R2) gain of chr8q was detected by day 258 and persisted (**Fig. 5B**) but was mutually exclusive with gains of chr3q in R1 and R3. In contrast, patterns of CNAs between different cultures from the same donor were more variable despite sharing some common CNAs (**Fig. 1C**).

DNA sequencing of ECBs from replicate cultures at regular intervals yields relative abundances of subclones over time. These data were used to construct Mueller plots to visualize clonal dynamics (as shown for 3 cultures in **Fig 5C**), where barcodes were assigned unique colors based on subclone frequencies across replicates within a culture and the highest frequency

subclone was colored red. For example, the red colored band in D2C1, replicate 1 represents the same barcoded subclone as in D2C1, replicates 2 and 3. For each culture (with the exception of D2C1R2) the same subclone became dominant across all replicates (**Fig. 5C**), consistent with an intrinsic fitness advantage and deterministic outgrowth (**Fig. 5A**). For D2C1 replicates R1 and R3 the red subclone became dominant, in-line with their shared copy number profiles. In contrast, for R2, the green subclone which acquired gain of chr8q (spanning the *MYC* oncogene), overtook the population. Intriguingly, the brown clone expanded concomitant with the green clone before going extinct, suggestive of their mutual dependence. For D2C2 the red barcode became dominant in all three replicates with similar dynamics amongst other subclones. For D3C2, by the late timepoint, several CNAs arose uniquely in individual replicates, including chr15q gain in R1, 11q gain in R2 and 6p loss in R3. Here, the red subclone rises to dominance most rapidly in R1, whereas expansion of this dominant clone is more gradual in R2 and R3. Of note, for R2 the barcode was lost ~day 273, presumably with the corresponding genomic segment, resulting in a shorter time course compared to R1 and R3.

The longitudinal WGS data indicate that CNA subclones co-existed for long periods of time (140 days for D3C1, 70 days for D2C2, **Fig 2C, S9**), but the increased resolution of the barcode data unveils competition over longer time frames. For example, in D2C2R2 and R3 the brown and blue subclones co-existed with the winning red subclone for the full time course (126-315d). Of note, the correlation in subclone frequencies across replicate experiments over time was generally high, suggesting similar population dynamics across all subclones, not only the one that eventually swept through the population (**Fig. S15D**). Especially striking is the remergence of the blue and purple subclones in D2C2 R2 and R3 at the mid-timepoint. Similar clonal dynamics were observed in two additional cultures (D1C1 and D2C3, **Fig. S15A-B**). Moreover, by constructing subclone specific growth curves and estimating their derivatives over time, we show that the winning subclone had high initial fitness but also increased in proliferative capacity with time (**Fig. 5D, Methods**). These patterns may reflect a 'rich-get-richer' effect where fitness advantages acquired early drive clonal expansions, thereby increasing the chances of acquiring future alterations that fuel growth (*2*).

In sum, lineage tracing in *TP53*$^{-/-}$ gastric organoids reveals highly reproducible dynamics across replicate cultures and broadly similar patterns across experiments with adaptive lineages sweeping rapidly to fixation (**Table S13**). For example, the dominant clones comprised 75% of the population by day 144 (median across cultures) after ECB transduction. These patterns are reminiscent of rapid adaptation observed in isogenic microbial populations (*36*, *50*) and attributable to standing variation which enhances short-term repeatability, due to adaptive mutations present in the initial population (*51*). They are also consistent with mounting evidence that even complex evolutionary processes have repeatable characteristics (*52*).

Genotype to phenotype mapping defines molecular determinants of winning subclones

The barcode frequency trajectories detailed above reveal the clonal dynamics of adaptation following *TP53* loss at exquisite resolution over year-long timescales. When coupled with longitudinal WGS, this enabled the identification of CNAs associated with fitness differences, albeit at a coarse-grain level. To further investigate the targets of selection and how they change through time and across populations, we leveraged the ECBs to jointly capture lineage and transcriptional states in individual cells. Moreover, by inferring copy number states from the single cell RNA-seq data, it is possible to define CNAs and expression programs associated with clonal dominance and extinction.

Specifically, we sought to characterize the molecular features of ‘winning’ subclones that dominated the population after prolonged evolution. To this end, we performed 10X scRNA-seq for several donors and replicates at selected timepoints when the population was heterogeneous, as determined based on barcode frequencies (**Fig. 5C**). For D2C2R2, which was sampled at day 173, 1,284 cells passed QC and we identified 20 subclones each with more than 10 cells, all of

which were among the top 38 most frequent ECBs based on barcode sequencing. Arm-level CNAs were inferred from the scRNA-seq data using *inferCNV* (**Methods**), revealing numerous subclone-specific CNAs (**Fig. 5B, 6A, Methods**). Nonetheless, the aggregate CNA landscapes were concordant with WGS data. Moreover, single cell DNA-sequencing of this sample on the 10x Chromium platform indicated similar profiles and comparable frequencies to the inferred subclone-specific CNAs based on scRNA-seq (**Fig. S17, Methods**).

A detailed examination of barcode frequencies and CNAs from this replicate (D2C2R2) reveals complex evolutionary dynamics amongst multiple coexisting subclones. Most cells comprising the winning subclone (ECB-0, red) acquired chromosome $3p^-$, $3q^+$, $9p^-$ and $9q^+$ alterations early, as these events were clonal or nearly clonal in the parent population at day 143 (**Fig. 6A,B**). A subpopulation within ECB-0 (termed "0a") acquired additional $chr4q^-$ and $chr20q^+$ alterations and ultimately became dominant with these alterations present in ~90% of the population at the latest time point (day 315) (**Fig. 5B, S18A**). Similar dynamics were seen across all replicate cultures where the winning subclones contained a nested CNA-defined subclone (which arise sequentially such that all cells harboring the second CNA also harbor the first) (**Fig. S18B-C, S19-21**). These data are in-line with the concept that *TP53* alterations distribute fitness over a large number of clones harboring distinct CNAs (*53*).

Since successful subclones consistently acquired additional genetic diversity in the form of CNAs, we sought to investigate the potential functional relevance of these events, focusing on a subset of subclones with divergent copy number. As an example, D2C2R2 consisted of at least five different CNA clones at the time of barcode insertion (**Fig. 6C, Table S14**). Multiple instances of convergent evolution were evident *within* this culture, where subclones acquired the same CNA independently, implying stringent selection for the target genes/locus. For example, ECB-0a, ECB-11 and ECB-56 each lost variable sized regions of chr4q. ECB-9 lacked common early alterations including $chr3p^-$, but subsequently acquired chr9p/q alterations. Despite the incomplete set of CNAs in ECB-9, its growth closely trailed that of the winning subclone (ECB-0, **Fig. 6D**). Convergent copy number evolution was also evident *across* cultures, where chr15 and chr20 amplifications were present in the majority of cells in D3C2 R1 and R2 at the latest time point, while these events plus chr11 amplification were present in an R3 subclone by the mid time point. Similar to the longitudinal WGS data, the single cell copy number profiles and lineage dynamics signal clonal interference in which beneficial alterations in separate subclones compete, while beneficial alterations in nested subclones reinforce one another (*52*).

Although genomic sequence affords a staggering number of evolutionary paths, genomic alterations map onto similar phenotypic effects. We therefore reasoned that highly fit subclones would share transcriptional programs and that such phenotypic convergence would be more constrained than at the genotypic level. For D2C2R2, the winning subclone 0a, but not its parent 0b, exhibited high expression of several well-known GC genes, including *CEACAM5*, *CEACAM6*, *CLDN3*, *CLDN4* and *CLDN7* (**Fig. 6E**). These genes were also more highly expressed in winning subclones of all other replicate cultures, except for ECB-1a (green) in D2C1R2 which acquired 8q gain (**Fig. S19D, S20D, S21D**). Other genes up-regulated in the winning subclone 0a (versus 0b) include *RNF186, MUC13, CCL20,* and *LGALS1* (**Fig. 6F, Table S15-S16**). The E3-ligase ring finger protein, *RNF186,* regulates intestinal homeostasis and is associated with ulcerative colitis (*54*). *MUC13* encodes a transmembrane mucin glycoprotein expressed on the apical surface of mucosal cells epithelial cells and is upregulated in intestinal metaplasia and GC (*55*, *56*). *CCL20* is a chemokine secreted by intestinal epithelia that mediates mucosal adaptive immune responses (*57*). *LGALS1* (galectin-1) promotes epithelial-mesenchymal transition, invasion and vascular mimicry in GC (*58*).

GSEA analysis, comparing the winning subclone in D2C2 to all other cells, extended these findings, revealing up-regulation of several pathways, including TNF-α signaling via NF-$\kappa\beta$, as well as hypoxia, apoptosis and p53 (**Fig. 6G, S22A, Table S17**). These same pathways were upregulated in three barcoded replicates for D2C2 at the final time point (day 315), as well as in

the non-barcoded D2C2 culture at *Mid* and *Late* time points (relative to *Early*) (**Fig. 6F**, right) and other donors/cultures (D1C1, D1C2, D1C3, D2C3, D3C2) (**Fig.3J**). Further, these pathways were upregulated in the winning subclone from multiple independent barcoded donors and cultures (**Fig. S19-21**), including the divergent subclone (ECB-1a, green) in D2C1R2 (**Fig. 5C**, **S20F**). Additionally, there was a significant difference in pathway activation between *Late* (relative to *Early*) and winning subclones, compared with all other analyzed subclones for p53, apoptosis and TNF-α signaling via NF-κβ pathways (Fisher exact test) (**Fig. S22B**). More generally, a high degree of correlation was seen amongst winning and late subclones across the top 10 altered gene sets. Notably, the winning subclones cluster with *Late* cultures (D1C1, D2C2, D2C3 and D3C2) that exhibited malignant transcriptional states based on the unsupervised LSI projection but not with those that had a WT like mucinous phenotype (D1C2 and D1C3) (**Fig. 6H, S22A, S23**). Together, these data highlight strong convergent phenotypic evolution in which the early (~100-200 days) activation of specific pathways is selectively advantageous, canalizing cells towards malignancy.

Discussion

We perform multi-year experimental evolution in *TP53* deficient gastric organoids to define causal relationships between this common initiating genetic lesion, cellular phenotypes and fitness. Remarkably, *TP53* KO alone recapitulates multiple features of the CIN subgroup of GC, including aneuploidy, specific CNAs, SVs and transcriptional programs, emphasizing the importance of cell intrinsic processes during pre-malignant evolution. Although aneuploidy is a substrate for evolution and propagates tumor heterogeneity (*21*), these data reveal preferred orders in the acquisition of CNAs with early loss of chr3p and 9p, frequently followed by bi-allelic inactivation of *CDKN2A* and/or *FHIT* and relatively late gain of 20q. Such preferred mutational orders have been described during tumorigenesis, most notably in the colon but the resolution of inferences from cross-sectional data or established tumors is limited (*10*, *59*). Our findings extend this principle to a tissue where the genomic features of precursor lesions are poorly understood and dominated by CNAs rather than point mutations.

While copy number evolution remains poorly understood, recent studies have begun to shed light on this process. For example, single cell sequencing of triple negative breast cancers indicates that CNAs can be acquired early (relative to tumor expansion) and in a punctuated fashion (*60*). Colorectal adenomas appear to traverse a rugged CNA fitness landscape and undergo a bottleneck during the transition to invasive disease before attaining quasi-stable karyotypes (*61*). The extensive clonal interference observed in our multi-year evolved gastric organoids suggests that they have not yet achieved an optimal karyotype. This is not surprising given that we are modeling the earliest stages of tumorigenesis, whereas malignant transformation and tumor expansion often occurs years after the first 'hit' (*9*, *10*). Consistent with this, evolved *TP53* deficient organoids capture many but not all features of invasive disease. For example, the CIN subgroup is enriched for *ERBB2* amplification and increased ploidy (*28*), neither of which were observed in our evolved organoids, presumably because they reflect pre-malignant states. This is consistent with the later acquisition of these events, including in a case-cohort study in BE where the co-selection of alterations on different chromosomes occurred 24-48 months prior to diagnosis of ESCA, whereas WGD was only detected in the subsequent cancer (*31*).

Despite harboring extensive genomic insults, the evolved *TP53* deficient gastric organoids lacked morphologic changes as has similarly been noted in human tissues, suggesting that these models not only provide a window into pre-neoplasia but mirror the latency of tumor initiation (*12*, *62*). Indeed, the long interval between intestinal metaplasia and GC presents opportunities for earlier intervention and improved risk stratification but necessitates an understanding of the molecular determinants that underpin progression (*63*). The quest to uncover these risk factors is complicated by the fact that GC comprises four distinct subtypes found in different anatomical locations at varied frequencies. While there are evidently distinct paths to malignant

transformation across these subgroups, we model the prevalent CIN subgroup which commonly arises in the body and gastroesophageal junction/cardia. Gastric cardia is also the presumed cell of origin for BE (*64*) which often exhibit *TP53* and *CDKN2A* loss and where CNAs predict progression to invasion (*65*). In light of our finding that *TP53* deficiency elicits CNAs in a defined temporal order, this raises the possibility that CNAs may similarly predict progression to CIN GC. Future evaluation of this hypothesis will require large, clinically and anatomically annotated, longitudinal tissue cohorts with long-term follow-up.

Given the lack of direct observations of human tumor initiation and growth, evolutionary dynamics have previously been inferred computationally from tumor genome sequencing data at a single or few timepoints (*5*, *6*, *66*). In particular, we previously inferred stringent positive selection from multi-region sequencing of BE (*6*), whereas matched adenocarcinomas exhibited patterns indicative of effectively neutral evolution, attributable to rapid growth after malignant transformation (*5*, *66*). These observations are in agreement with the empirical data presented here based on time resolved WGS and prospective lineage tracing which revealed stringent selection during pre-neoplastic evolution and reproducible subclonal dynamics across replicate *TP53* deficient gastric organoids cultures in which the same, initially rare, subclone fixed in the population. The dense temporal resolution afforded by this experimental platform also exposed pervasive clonal interference amongst CNA clones, accompanied by intermittent periods of relative stasis. By jointly capturing lineage, CNAs and transcriptional states in individual cells, we investigated the molecular basis for clonal expansion and persistence, identifying phenotypic convergence on common pathways, including TNF-α and hypoxic signaling in the dominant subclone across cultures and donors, despite divergent copy number profiles. Strong adaptive pressure generates convergent evolution as evident in this system at both the genomic copy number and pathway levels, where the latter was stronger and suggests a surprising degree of predictability.

In theory, evolution is predictable to the extent that both positive and negative selection canalizes adaptation towards a dominant path (*52*), but this is all the more notable given the starting point of these experiments was genetically heterogeneous gastric tissue. Indeed, it is now appreciated that post-zygotic tissues are somatic mosaic as a result of age-related mutation accumulation (*7*). Additionally, obesity is risk factor for GC associated with inflammation (*39*), which may influence cell phenotypes and clonal composition in the biopsies used to establish the derivative organoid lines studied here. These models thus embed human biology relevant to tumorigenesis that are not captured by induced pluripotent stem cells or murine organoids.

Our work provides proof of concept that experimental evolution in cancer driver-engineered epithelial organoids enables the unbiased identification of selectively advantageous alterations and constraints in the order in which they accumulate. Such constraints, due to epistasis, can reveal barriers to malignant transformation and nominate potential therapeutic targets. Understanding pre-neoplastic processes under constant conditions represents a critical starting point, but future studies could expand on this by modulating the environment or introducing *H. pylori*, a GC risk factor. More generally, while the space of possible initiating insults is vast, recurrent tissue-specific alterations nominated through sequencing studies can similarly be modeled in other tissues individually and in combination. Much as the long-term experimental evolution experiments pioneered by Lenski and colleagues two decades ago continue to yield fundamental insights into microbial adaptation (*3*, *4*), we anticipate that our results will advance empirical and theoretical investigations of selection, mutation and genome instability during clonal evolution in human cells.

Methods

A detailed description of the Materials and Methods is available in the Supplementary Materials.

References

1. I. Vázquez-García, F. Salinas, J. Li, A. Fischer, B. Barré, J. Hallin, A. Bergström, E. Alonso-Perez, J. Warringer, V. Mustonen, G. Liti, Clonal Heterogeneity Influences the Fate of New Adaptive Mutations. *Cell Rep.* **21**, 732–744 (2017).
2. A. N. Nguyen Ba, I. Cvijović, J. I. Rojas Echenique, K. R. Lawrence, A. Rego-Costa, X. Liu, S. F. Levy, M. M. Desai, High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature*. **575**, 494–499 (2019).
3. R. E. Lenski, M. R. Rose, S. C. Simpson, S. C. Tadler, Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. *Am. Nat.* **138**, 1315–1341 (1991).
4. B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, M. M. Desai, The dynamics of molecular evolution over 60,000 generations. *Nature*. **551**, 45–50 (2017).
5. A. Sottoriva, H. Kang, Z. Ma, T. A. Graham, M. P. Salomon, J. Zhao, P. Marjoram, K. Siegmund, M. F. Press, D. Shibata, C. Curtis, A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015).
6. R. Sun, Z. Hu, A. Sottoriva, T. A. Graham, A. Harpak, Z. Ma, J. M. Fischer, D. Shibata, C. Curtis, Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nat. Genet.* **49**, 1015–1024 (2017).
7. E. Laconi, F. Marongiu, J. DeGregori, Cancer as a disease of old age: changing mutational and microenvironmental landscapes. *Br. J. Cancer*. **122**, 943–952 (2020).
8. J. L. Tsao, Y. Yatabe, R. Salovaara, H. J. Jarvinen, J. P. Mecklin, L. A. Aaltonen, S. Tavare, D. Shibata, Genetic reconstruction of individual colorectal tumor histories. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1236–1241 (2000).
9. A.-M. Baker, T. A. Graham, N. A. Wright, Pre-tumour clones, periodic selection and clonal interference in the origin and progression of gastrointestinal cancer: potential for biomarker development. *J. Pathol.* **229**, 502–514 (2013).
10. M. Gerstung, C. Jolly, I. Leshchiner, S. C. Dentro, S. Gonzalez, D. Rosebrock, T. J. Mitchell, Y. Rubanova, P. Anur, K. Yu, M. Tarabichi, A. Deshwar, J. Wintersinger, K. Kleinheinz, I. Vázquez-García, K. Haase, L. Jerman, S. Sengupta, G. Macintyre, S. Malikic, N. Donmez, D. G. Livitz, M. Cmero, J. Demeulemeester, S. Schumacher, Y. Fan, X. Yao, J. Lee, M. Schlesner, P. C. Boutros, D. D. Bowtell, H. Zhu, G. Getz, M. Imielinski, R. Beroukhim, S. C. Sahinalp, Y. Ji, M. Peifer, F. Markowetz, V. Mustonen, K. Yuan, W. Wang, Q. D. Morris, PCAWG Evolution & Heterogeneity Working Group, P. T. Spellman, D. C. Wedge, P. Van Loo, PCAWG Consortium, The evolutionary history of 2,658 cancers. *Nature*. **578**, 122–128 (2020).
11. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. **71**, 209–249 (2021).
12. W. Waddingham, S. A. V. Nieuwenburg, S. Carlson, M. Rodriguez-Justo, M. Spaander, E. J. Kuipers, M. Jansen, D. G. Graham, M. Banks, Recent advances in the detection and management of early gastric cancer and its precursors. *Frontline Gastroenterol.* **12**, 322–331 (2021).
13. Y.-H. Lo, K. S. Kolahi, Y. Du, C.-Y. Chang, A. Krokhotin, A. Nair, W. D. Sobba, K. Karlsson, S. J. Jones, T. A. Longacre, A. T. Mah, B. Tercan, A. Sockell, H. Xu, J. A. Seoane, J. Chen, I. Shmulevich, J. S. Weissman, C. Curtis, A. Califano, H. Fu, G. R. Crabtree, C. J. Kuo, A CRISPR/Cas9-Engineered ARID1A-Deficient Human Gastric Cancer Organoid Model Reveals Essential and Nonessential Modes of Oncogenic Transformation. *Cancer Discov.* **11**, 1562–1581 (2021).
14. N. S. Sethi, O. Kikuchi, G. N. Duronio, M. D. Stachler, J. M. McFarland, R. Ferrer-Luna, Y. Zhang, C. Bao, R. Bronson, D. Patil, F. Sanchez-Vega, J.-B. Liu, E. Sicinska, J.-B. Lazaro,

K. L. Ligon, R. Beroukhim, A. J. Bass, Early TP53 alterations engage environmental exposures to promote gastric premalignancy in an integrative mouse model. *Nat. Genet.* **52**, 219–230 (2020).
15. T. Seidlitz, B.-K. Koo, D. E. Stange, Gastric organoids-an in vitro model system for the study of gastric development and road to personalized medicine. *Cell Death Differ.* **28**, 68–83 (2021).
16. H. H. N. Yan, H. C. Siu, S. Law, S. L. Ho, S. S. K. Yue, W. Y. Tsui, D. Chan, A. S. Chan, S. Ma, K. O. Lam, S. Bartfeld, A. H. Y. Man, B. C. H. Lee, A. S. Y. Chan, J. W. H. Wong, P. S. W. Cheng, A. K. W. Chan, J. Zhang, J. Shi, X. Fan, D. L. W. Kwong, T. W. Mak, S. T. Yuen, H. Clevers, S. Y. Leung, A Comprehensive Human Gastric Cancer Organoid Biobank Captures Tumor Subtype Heterogeneity and Enables Therapeutic Screening. *Cell Stem Cell*. **23**, 882-897.e11 (2018).
17. Y.-H. Lo, K. Karlsson, C. J. Kuo, Applications of Organoids for Cancer Biology and Precision Medicine. *Nat Cancer*. **1**, 761–773 (2020).
18. A. M. Taylor, J. Shih, G. Ha, G. F. Gao, X. Zhang, A. C. Berger, S. E. Schumacher, C. Wang, H. Hu, J. Liu, A. J. Lazar, Cancer Genome Atlas Research Network, A. D. Cherniack, R. Beroukhim, M. Meyerson, Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell*. **33**, 676-689.e3 (2018).
19. S. L. Thompson, D. A. Compton, Examining the link between chromosomal instability and aneuploidy in human cells. *J. Cell Biol.* **180**, 665–672 (2008).
20. L. J. Valente, A. Tarangelo, A. M. Li, M. Naciri, N. Raj, A. M. Boutelle, Y. Li, S. S. Mello, K. Bieging-Rolett, R. J. DeBerardinis, J. Ye, S. J. Dixon, L. D. Attardi, p53 deficiency triggers dysregulation of diverse cellular processes in physiological oxygen. *J. Cell Biol.* **219** (2020), doi:10.1083/jcb.201908212.
21. U. Ben-David, A. Amon, Context is everything: aneuploidy in cancer. *Nat. Rev. Genet.* **21**, 44–62 (2020).
22. A. Narkar, B. A. Johnson, P. Bharne, J. Zhu, V. Padmanaban, D. Biswas, A. Fraser, P. A. Iglesias, A. J. Ewald, R. Li, On the role of p53 in the cellular response to aneuploidy. *Cell Rep.* **34**, 108892 (2021).
23. M. B. Weiss, M. I. Vitolo, M. Mohseni, D. M. Rosen, S. R. Denmeade, B. H. Park, D. J. Weber, K. E. Bachman, Deletion of p53 in human mammary epithelial cells causes chromosomal instability and altered therapeutic response. *Oncogene*. **29**, 4715–4724 (2010).
24. L. N. Redman-Rivera, T. M. Shaver, H. Jin, C. B. Marshall, J. M. Schafer, Q. Sheng, R. A. Hongo, K. E. Beckermann, F. C. Wheeler, B. D. Lehmann, J. A. Pietenpol, Acquisition of aneuploidy drives mutant p53-associated gain-of-function phenotypes. *Nat. Commun.* **12**, 5184 (2021).
25. J. Drost, R. H. van Jaarsveld, B. Ponsioen, C. Zimberlin, R. van Boxtel, A. Buijs, N. Sachs, R. M. Overmeer, G. J. Offerhaus, H. Begthel, J. Korving, M. van de Wetering, G. Schwank, M. Logtenberg, E. Cuppen, H. J. Snippert, J. P. Medema, G. J. P. L. Kops, H. Clevers, Sequential cancer mutations in cultured human intestinal stem cells. *Nature*. **521**, 43–47 (2015).
26. E. R. Kastenhuber, S. W. Lowe, Putting p53 in Context. *Cell*. **170**, 1062–1078 (2017).
27. A. M. Boutelle, L. D. Attardi, p53 and Tumor Suppression: It Takes a Network. *Trends Cell Biol.* **31**, 298–310 (2021).
28. Cancer Genome Atlas Research Network, Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. **513**, 202–209 (2014).
29. M. T. Barrett, C. A. Sanchez, L. J. Prevo, D. J. Wong, P. C. Galipeau, T. G. Paulson, P. S. Rabinovitch, B. J. Reid, Evolution of neoplastic cell lineages in Barrett oesophagus. *Nat. Genet.* **22**, 106–109 (1999).

30. J. C. Saldivar, D. Park, Mechanisms shaping the mutational landscape of the FRA3B/FHIT-deficient cancer genome. *Genes Chromosomes Cancer*. **58**, 317–323 (2019).
31. X. Li, P. C. Galipeau, T. G. Paulson, C. A. Sanchez, J. Arnaudo, K. Liu, C. L. Sather, R. L. Kostadinov, R. D. Odze, M. K. Kuhner, C. C. Maley, S. G. Self, T. L. Vaughan, P. L. Blount, B. J. Reid, Temporal and spatial evolution of somatic chromosomal alterations: a case-cohort study of Barrett's esophagus. *Cancer Prev. Res.* **7**, 114–127 (2014).
32. F. Newell, K. Patel, M. Gartside, L. Krause, S. Brosda, L. G. Aoude, K. A. Loffler, V. F. Bonazzi, A.-M. Patch, S. H. Kazakoff, O. Holmes, Q. Xu, S. Wood, C. Leonard, G. Lampe, R. V. Lord, D. C. Whiteman, J. V. Pearson, K. Nones, N. Waddell, A. P. Barbour, Complex structural rearrangements are present in high-grade dysplastic Barrett's oesophagus samples. *BMC Med. Genomics*. **12**, 31 (2019).
33. L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, S. A. J. R. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A.-L. Børresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjörd, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jäger, D. T. W. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. López-Otín, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. J. Tutt, R. Valdés-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, M. R. Stratton, Signatures of mutational processes in human cancer. *Nature*. **500**, 415–421 (2013).
34. K. Hadi, X. Yao, J. M. Behr, A. Deshpande, C. Xanthopoulakis, H. Tian, S. Kudman, J. Rosiene, M. Darmofal, J. DeRose, R. Mortensen, E. M. Adney, A. Shaiber, Z. Gajic, M. Sigouros, K. Eng, J. A. Wala, K. O. Wrzeszczyński, K. Arora, M. Shah, A. K. Emde, V. Felice, M. O. Frank, R. B. Darnell, M. Ghandi, F. Huang, S. Dewhurst, J. Maciejowski, T. de Lange, J. Setton, N. Riaz, J. S. Reis-Filho, S. Powell, D. A. Knowles, E. Reznik, B. Mishra, R. Beroukhim, M. C. Zody, N. Robine, K. M. Oman, C. A. Sanchez, M. K. Kuhner, L. P. Smith, P. C. Galipeau, T. G. Paulson, B. J. Reid, X. Li, D. Wilkes, A. Sboner, J. M. Mosquera, O. Elemento, M. Imielinski, Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. *Cell*. **183**, 197-210.e32 (2020).
35. T. W. Glover, T. E. Wilson, M. F. Arlt, Fragile sites in cancer: more than meets the eye. *Nat. Rev. Cancer*. **17**, 489–501 (2017).
36. G. I. Lang, D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein, M. M. Desai, Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*. **500**, 571–574 (2013).
37. C. Bock, M. Boutros, J. G. Camp, L. Clarke, H. Clevers, J. A. Knoblich, P. Liberali, A. Regev, A. C. Rios, O. Stegle, H. G. Stunnenberg, S. A. Teichmann, B. Treutlein, R. G. J. Vries, Human Cell Atlas 'Biological Network' Organoids, The Organoid Cell Atlas. *Nat. Biotechnol.* **39**, 13–17 (2021).
38. G. M. H. Birchenough, M. EV Johansson, J. K. Gustafsson, J. H. Bergström, G. C. Hansson, New developments in goblet cell mucus secretion and function. *Mucosal Immunology*. **8** (2015), pp. 712–719.
39. M. Camilleri, H. Malhi, A. Acosta, Gastrointestinal Complications of Obesity. *Gastroenterology*. **152**, 1656–1670 (2017).
40. S. Reber, J. Mechtersheimer, S. Nasif, J. A. Benitez, M. Colombo, M. Domanski, D. Jutzi, E. Hedlund, M.-D. Ruepp, CRISPR-Trap: a clean approach for the generation of gene knockouts and gene replacements in human cells. *Mol. Biol. Cell*. **29**, 75–83 (2018).

41. D. Dong, Z. Mu, C. Zhao, M. Sun, ZFAS1: a novel tumor-related long non-coding RNA. *Cancer Cell Int.* **18**, 125 (2018).
42. S. Rao, X. Yang, K. Ohshiro, S. Zaidi, Z. Wang, K. Shetty, X. Xiang, M. I. Hassan, T. Mohammad, P. S. Latham, B.-N. Nguyen, L. Wong, H. Yu, Y. Al-Abed, B. Mishra, M. Vacca, G. Guenigault, M. E. D. Allison, A. Vidal-Puig, J. N. Benhammou, M. Alvarez, P. Pajukanta, J. R. Pisegna, L. Mishra, β2-spectrin (SPTBN1) as a therapeutic target for diet-induced liver disease and preventing cancer development. *Sci. Transl. Med.* **13**, eabk2267 (2021).
43. M. Zhang, S. Hu, M. Min, Y. Ni, Z. Lu, X. Sun, J. Wu, B. Liu, X. Ying, Y. Liu, Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. *Gut*. **70**, 464–475 (2021).
44. G. A. Webster, N. D. Perkins, Transcriptional cross talk between NF-kappaB and p53. *Mol. Cell. Biol.* **19**, 3485–3495 (1999).
45. N. Sethi, O. Kikuchi, J. McFarland, Y. Zhang, M. Chung, N. Kafker, M. Islam, B. Lampson, A. Chakraborty, W. G. Kaelin Jr, A. J. Bass, Mutant p53 induces a hypoxia transcriptional program in gastric and esophageal adenocarcinoma. *JCI Insight*. **4** (2019), doi:10.1172/jci.insight.128439.
46. J. M. Granja, S. Klemm, L. M. McGinnis, A. S. Kathiria, A. Mezger, M. R. Corces, B. Parks, E. Gars, M. Liedtke, G. X. Y. Zheng, H. Y. Chang, R. Majeti, W. J. Greenleaf, Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **37**, 1458–1465 (2019).
47. M. Andreatta, J. Corria-Osorio, S. Müller, R. Cubas, G. Coukos, S. J. Carmona, Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat. Commun.* **12**, 2965 (2021).
48. A. Sathe, S. M. Grimes, B. T. Lau, J. Chen, C. Suarez, R. J. Huang, G. Poultsides, H. P. Ji, Single-Cell Genomic Characterization Reveals the Cellular Reprogramming of the Gastric Tumor Microenvironment. *Clinical Cancer Research*. **26** (2020), pp. 2640–2653.
49. P. Zhang, M. Yang, Y. Zhang, S. Xiao, X. Lai, A. Tan, S. Du, S. Li, Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer. *Cell Rep.* **30**, 4317 (2020).
50. S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, G. Sherlock, Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*. **519**, 181–186 (2015).
51. M. K. Burke, G. Liti, A. D. Long, Standing genetic variation drives repeatable experimental evolution in outcrossing populations of Saccharomyces cerevisiae. *Mol. Biol. Evol.* **31**, 3228–3239 (2014).
52. M. Lassig, V. Mustonen, A. M. Walczak, Predicting evolution. *Nat Ecol Evol*. **1**, 77 (2017).
53. S. Salehi, F. Kabeer, N. Ceglia, M. Andronescu, M. J. Williams, K. R. Campbell, T. Masud, B. Wang, J. Biele, J. Brimhall, D. Gee, H. Lee, J. Ting, A. W. Zhang, H. Tran, C. O'Flanagan, F. Dorri, N. Rusk, T. R. de Algara, S. R. Lee, B. Y. C. Cheng, P. Eirew, T. Kono, J. Pham, D. Grewal, D. Lai, R. Moore, A. J. Mungall, M. A. Marra, IMAXT Consortium, A. McPherson, A. Bouchard-Côté, S. Aparicio, S. P. Shah, Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature*. **595**, 585–590 (2021).
54. K. Fujimoto, M. Kinoshita, H. Tanaka, D. Okuzaki, Y. Shimada, H. Kayama, R. Okumura, Y. Furuta, M. Narazaki, A. Tamura, S. Hatakeyama, M. Ikawa, K. Tsuchiya, M. Watanabe, A. Kumanogoh, S. Tsukita, K. Takeda, Regulation of intestinal homeostasis by the ulcerative colitis-associated gene RNF186. *Mucosal Immunol.* **10**, 446–459 (2017).
55. Y. H. Sheng, S. Triyana, R. Wang, I. Das, K. Gerloff, T. H. Florin, P. Sutton, M. A. McGuckin, MUC1 and MUC13 differentially regulate epithelial inflammation in response to inflammatory and infectious stimuli. *Mucosal Immunol.* **6**, 557–568 (2013).

56. A. Takeno, I. Takemasa, S. Seno, M. Yamasaki, M. Motoori, H. Miyata, K. Nakajima, S. Takiguchi, Y. Fujiwara, T. Nishida, T. Okayama, K. Matsubara, Y. Takenaka, H. Matsuda, M. Monden, M. Mori, Y. Doki, Gene expression profile prospectively predicts peritoneal relapse after curative surgery of gastric cancer. *Ann. Surg. Oncol.* **17**, 1033–1042 (2010).
57. A. Izadpanah, M. B. Dwinell, L. Eckmann, N. M. Varki, M. F. Kagnoff, Regulated MIP-3alpha/CCL20 production by human intestinal epithelium: mechanism for modulating mucosal immunity. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280**, G710-9 (2001).
58. X. You, J. Wu, Y. Wang, Q. Liu, Z. Cheng, X. Zhao, G. Liu, C. Huang, J. Dai, Y. Zhou, D. Chen, Y. Chong, Galectin-1 promotes vasculogenic mimicry in gastric adenocarcinoma via the Hedgehog/GLI signaling pathway. *Aging* . **12**, 21837–21853 (2020).
59. E. R. Fearon, B. Vogelstein, A genetic model for colorectal tumorigenesis. *Cell*. **61**, 759–767 (1990).
60. R. Gao, A. Davis, T. O. McDonald, E. Sei, X. Shi, Y. Wang, P. C. Tsai, A. Casasent, J. Waters, H. Zhang, F. Meric-Bernstam, F. Michor, N. E. Navin, Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat. Genet.* **48**, 1119–1130 (2016).
61. W. Cross, M. Kovac, V. Mustonen, D. Temko, H. Davis, A. M. Baker, S. Biswas, R. Arnold, L. Chegwidden, C. Gatenbee, A. R. Anderson, V. H. Koelzer, P. Martinez, X. Jiang, E. Domingo, D. J. Woodcock, Y. Feng, M. Kovacova, T. Maughan, S. Cort Consortium, M. Jansen, M. Rodriguez-Justo, S. Ashraf, R. Guy, C. Cunningham, J. E. East, D. C. Wedge, L. M. Wang, C. Palles, K. Heinimann, A. Sottoriva, S. J. Leedham, T. A. Graham, I. P. M. Tomlinson, The evolutionary landscape of colorectal tumorigenesis. *Nat Ecol Evol*. **2**, 1661–1672 (2018).
62. M. J. Whitson, G. W. Falk, Predictors of progression to high-grade dysplasia or adenocarcinoma in Barrett's esophagus. *Gastroenterol. Clin. North Am.* **44**, 299–315 (2015).
63. K. K. Huang, K. Ramnarayanan, F. Zhu, S. Srivastava, C. Xu, A. L. K. Tan, M. Lee, S. Tay, K. Das, M. Xing, A. Fatehullah, S. M. F. Alkaff, T. K. H. Lim, J. Lee, K. Y. Ho, S. G. Rozen, B. T. Teh, N. Barker, C. K. Chia, C. Khor, C. J. Ooi, K. M. Fock, J. So, W. C. Lim, K. L. Ling, T. L. Ang, A. Wong, J. Rao, A. Rajnakova, L. G. Lim, W. M. Yap, M. Teh, K. G. Yeoh, P. Tan, Genomic and Epigenomic Profiling of High-Risk Intestinal Metaplasia Reveals Molecular Determinants of Progression to Gastric Cancer. *Cancer Cell*. **33**, 137-150.e5 (2018).
64. K. Nowicki-Osuch, L. Zhuang, S. Jammula, C. W. Bleaney, K. T. Mahbubani, G. Devonshire, A. Katz-Summercorn, N. Eling, A. Wilbrey-Clark, E. Madissoon, J. Gamble, M. Di Pietro, M. O'Donovan, K. B. Meyer, K. Saeb-Parsy, A. D. Sharrocks, S. A. Teichmann, J. C. Marioni, R. C. Fitzgerald, Molecular phenotyping reveals the identity of Barrett's esophagus and its malignant transition. *Science*. **373**, 760–767 (2021).
65. S. Killcoyne, E. Gregson, D. C. Wedge, D. J. Woodcock, M. D. Eldridge, R. de la Rue, A. Miremadi, S. Abbas, A. Blasko, C. Kosmidou, W. Januszewicz, A. V. Jenkins, M. Gerstung, R. C. Fitzgerald, Genomic copy number predicts esophageal cancer years before transformation. *Nat. Med.* **26**, 1726–1732 (2020).
66. M. J. Williams, B. Werner, T. Heide, C. Curtis, C. P. Barnes, A. Sottoriva, T. A. Graham, Quantification of subclonal selection in cancer from bulk sequencing data. *Nat. Genet.* **50**, 895–903 (2018).

Supplementary References

67. T. Sato, D. E. Stange, M. Ferrante, R. G. J. Vries, J. H. Van Es, S. Van den Brink, W. J. Van Houdt, A. Pronk, J. Van Gorp, P. D. Siersema, H. Clevers, Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology*. **141**, 1762–1772 (2011).

68. G. Schwank, B.-K. Koo, V. Sasselli, J. F. Dekkers, I. Heo, T. Demircan, N. Sasaki, S. Boymans, E. Cuppen, C. K. van der Ent, E. E. S. Nieuwenhuis, J. M. Beekman, H. Clevers, Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell*. **13**, 653–658 (2013).
69. S. F. Roerink, N. Sasaki, H. Lee-Six, M. D. Young, L. B. Alexandrov, S. Behjati, T. J. Mitchell, S. Grossmann, H. Lightfoot, D. A. Egan, A. Pronk, N. Smakman, J. van Gorp, E. Anderson, S. J. Gamble, C. Alder, M. van de Wetering, P. J. Campbell, M. R. Stratton, H. Clevers, Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature*. **556**, 457–462 (2018).
70. E. K. Brinkman, T. Chen, M. Amendola, B. van Steensel, Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
71. J. T. Neal, X. Li, J. Zhu, V. Giangarra, C. L. Grzeskowiak, J. Ju, I. H. Liu, S.-H. Chiou, A. A. Salahudeen, A. R. Smith, B. C. Deutsch, L. Liao, A. J. Zemek, F. Zhao, K. Karlsson, L. M. Schultz, T. J. Metzner, L. D. Nadauld, Y.-Y. Tseng, S. Alkhairy, C. Oh, P. Keskula, D. Mendoza-Villanueva, F. M. De La Vega, P. L. Kunz, J. C. Liao, J. T. Leppert, J. B. Sunwoo, C. Sabatti, J. S. Boehm, W. C. Hahn, G. X. Y. Zheng, M. M. Davis, C. J. Kuo, Organoid Modeling of the Tumor Immune Microenvironment. *Cell*. **175**, 1972-1988.e16 (2018).
72. H.-E. C. Bhang, D. A. Ruddy, V. Krishnamurthy Radhakrishna, J. X. Caushi, R. Zhao, M. M. Hims, A. P. Singh, I. Kao, D. Rakiec, P. Shaw, M. Balak, A. Raza, E. Ackley, N. Keen, M. R. Schlabach, M. Palmer, R. J. Leary, D. Y. Chiang, W. R. Sellers, F. Michor, V. G. Cooke, J. M. Korn, F. Stegmeier, Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat. Med.* **21**, 440–448 (2015).
73. M. Stoeckius, S. Zheng, B. Houck-Loomis, S. Hao, B. Z. Yeung, W. M. Mauck 3rd, P. Smibert, R. Satija, Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
74. M. Garcia, S. Juhos, M. Larsson, P. I. Olason, M. Martin, J. Eisfeldt, S. DiLorenzo, J. Sandgren, T. Díaz De Ståhl, P. Ewels, V. Wirta, M. Nistér, M. Käller, B. Nystedt, Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Res.* **9**, 63 (2020).
75. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).
76. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
77. I. Scheinin, D. Sie, H. Bengtsson, M. A. Van De Wiel, A. B. Olshen, H. F. Van Thuijl, H. F. Van Essen, P. P. Eijk, F. Rustenburg, G. A. Meijer, Others, DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* **24**, 2022–2032 (2014).
78. S. Kim, K. Scheffler, A. L. Halpern, M. A. Bekritsky, E. Noh, M. Källberg, X. Chen, Y. Kim, D. Beyter, P. Krusche, C. T. Saunders, Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods*. **15**, 591–594 (2018).
79. D. L. Cameron, J. Baber, C. Shale, A. T. Papenfuss, J. E. Valle-Inclan, N. Besselink, E. Cuppen, P. Priestley, GRIDSS, PURPLE, LINX: Unscrambling the tumor genome via integrated analysis of structural variation and copy number. *bioRxiv* (2019), p. 781013.
80. P. Priestley, J. Baber, M. P. Lolkema, N. Steeghs, E. de Bruijn, C. Shale, K. Duyvesteyn, S. Haidari, A. van Hoeck, W. Onstenk, P. Roepman, M. Voda, H. J. Bloemendal, V. C. G. Tjan-Heijnen, C. M. L. van Herpen, M. Labots, P. O. Witteveen, E. F. Smit, S. Sleijfer, E. E. Voest, E. Cuppen, Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature*. **575**, 210–216 (2019).

81. D. Endesfelder, R. Burrell, N. Kanu, N. McGranahan, M. Howell, P. J. Parker, J. Downward, C. Swanton, M. Kschischo, Chromosomal instability selects gene copy-number variants encoding core regulators of proliferation in ER+ breast cancer. *Cancer Res.* **74**, 4853–4863 (2014).
82. X. Chen, O. Schulz-Trieglaff, R. Shaw, B. Barnes, F. Schlesinger, M. Källberg, A. J. Cox, S. Kruglyak, C. T. Saunders, Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. **32**, 1220–1222 (2015).
83. T. Rausch, T. Zichner, A. Schlattl, A. M. Stütz, V. Benes, J. O. Korbel, DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. **28**, i333–i339 (2012).
84. J. A. Wala, P. Bandopadhayay, N. F. Greenwald, R. O'Rourke, T. Sharpe, C. Stewart, S. Schumacher, Y. Li, J. Weischenfeldt, X. Yao, C. Nusbaum, P. Campbell, G. Getz, M. Meyerson, C.-Z. Zhang, M. Imielinski, R. Beroukhim, SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* **28**, 581–591 (2018).
85. K. Wang, S. T. Yuen, J. Xu, S. P. Lee, H. H. N. Yan, S. T. Shi, H. C. Siu, S. Deng, K. M. Chu, S. Law, K. H. Chan, A. S. Y. Chan, W. Y. Tsui, S. L. Ho, A. K. W. Chan, J. L. K. Man, V. Foglizzo, M. K. Ng, A. S. Chan, Y. P. Ching, G. H. W. Cheng, T. Xie, J. Fernandez, V. S. W. Li, H. Clevers, P. A. Rejto, M. Mao, S. Y. Leung, Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **46**, 573–582 (2014).
86. G. Ha, A. Roth, J. Khattra, J. Ho, D. Yap, L. M. Prentice, N. Melnyk, A. McPherson, A. Bashashati, E. Laks, J. Biele, J. Ding, A. Le, J. Rosner, K. Shumansky, M. A. Marra, C. B. Gilks, D. G. Huntsman, J. N. McAlpine, S. Aparicio, S. P. Shah, TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome Res.* **24**, 1881–1893 (2014).
87. R. Shen, V. E. Seshan, FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
88. S. Zaccaria, B. J. Raphael, Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. *Nat. Commun.* **11**, 1–13 (2020).
89. S. Das, L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P.-R. Loh, W. G. Iacono, A. Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, C. Fuchsberger, Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
90. C. A. Miller, J. McMichael, H. X. Dang, C. A. Maher, L. Ding, T. J. Ley, E. R. Mardis, R. K. Wilson, Visualizing tumor evolution with the fishplot package for R. *BMC Genomics*. **17**, 880 (2016).
91. V. Obenchain, M. Lawrence, V. Carey, S. Gogarten, P. Shannon, M. Morgan, VariantAnnotation : a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics*. **30**, 2076–2078 (2014).
92. J. G. Tate, S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jupe, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell, S. A. Forbes, COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
93. L. B. Alexandrov, J. Kim, N. J. Haradhvala, M. N. Huang, A. W. Tian Ng, Y. Wu, A. Boot, K. R. Covington, D. A. Gordenin, E. N. Bergstrom, S. M. A. Islam, N. Lopez-Bigas, L. J. Klimczak, J. R. McPherson, S. Morganella, R. Sabarinathan, D. A. Wheeler, V. Mustonen, G. Getz, S. G. Rozen, M. R. Stratton, The repertoire of mutational signatures in human cancer. *Nature*. **578**, 94–101 (2020).

94. K. Gori, A. Baez-Ortega, sigfit: flexible Bayesian inference of mutational signatures. *bioRxiv* (2020), p. 372896.
95. T. Smith, A. Heger, I. Sudbery, UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
96. I. Tirosh, A. S. Venteicher, C. Hebert, L. E. Escalante, A. P. Patel, K. Yizhak, J. M. Fisher, C. Rodman, C. Mount, M. G. Filbin, C. Neftel, N. Desai, J. Nyman, B. Izar, C. C. Luo, J. M. Francis, A. A. Patel, M. L. Onozato, N. Riggi, K. J. Livak, D. Gennert, R. Satija, B. V. Nahed, W. T. Curry, R. L. Martuza, R. Mylvaganam, A. J. Iafrate, M. P. Frosch, T. R. Golub, M. N. Rivera, G. Getz, O. Rozenblatt-Rosen, D. P. Cahill, M. Monje, B. E. Bernstein, D. N. Louis, A. Regev, M. L. Suva, Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature*. **539**, 309–313 (2016).
97. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
98. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems*. **8**, 329-337.e4 (2019).
99. A. Sathe, S. M. Grimes, B. T. Lau, J. Chen, C. Suarez, R. J. Huang, G. Poultsides, H. P. Ji, Single-Cell Genomic Characterization Reveals the Cellular Reprogramming of the Gastric Tumor Microenvironment. *Clin. Cancer Res.* **26**, 2640–2653 (2020).
100. P. Zhang, M. Yang, Y. Zhang, S. Xiao, X. Lai, A. Tan, S. Du, S. Li, Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer. *Cell Rep.* **27**, 1934-1947.e5 (2019).
101. X. Han, Z. Zhou, L. Fei, H. Sun, R. Wang, Y. Chen, H. Chen, J. Wang, H. Tang, W. Ge, Y. Zhou, F. Ye, M. Jiang, J. Wu, Y. Xiao, X. Jia, T. Zhang, X. Ma, Q. Zhang, X. Bai, S. Lai, C. Yu, L. Zhu, R. Lin, Y. Gao, M. Wang, Y. Wu, J. Zhang, R. Zhan, S. Zhu, H. Hu, C. Wang, M. Chen, H. Huang, T. Liang, J. Chen, W. Wang, D. Zhang, G. Guo, Construction of a human cell landscape at single-cell level. *Nature*. **581**, 303–309 (2020).
102. S. Gao, L. Yan, R. Wang, J. Li, J. Yong, X. Zhou, Y. Wei, X. Wu, X. Wang, X. Fan, J. Yan, X. Zhi, Y. Gao, H. Guo, X. Jin, W. Wang, Y. Mao, F. Wang, L. Wen, W. Fu, H. Ge, J. Qiao, F. Tang, Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nat. Cell Biol.* **20**, 721–734 (2018).
103. Z. Tang, C. Li, B. Kang, G. Gao, C. Li, Z. Zhang, GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**, W98–W102 (2017).
104. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell* (2019), doi:10.1016/j.cell.2019.05.031.
105. T. Tickle, I. Tirosh, C. Georgescu, M. Brown, B. Haas, inferCNV of the Trinity CTAT Project. *Klarman Cell Observatory, Broad Institute of MIT and Harvard* (2019).
106. A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suvà, A. Regev, B. E. Bernstein, Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. **344** (2014), pp. 1396–1401.
107. J. Fan, H.-O. Lee, S. Lee, D.-E. Ryu, S. Lee, C. Xue, S. J. Kim, K. Kim, N. Barkas, P. J. Park, W.-Y. Park, P. V. Kharchenko, Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Res.* **28**, 1217–1227 (2018).
108. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. **1**, 417–425 (2015).

Acknowledgments
The authors thank Zheng Hu Susanne Tilk and Laura Attardi for helpful discussions, the Stanford University Hospital Tissue Procurement Shared Resource facility for specimen procurement and the Stanford Functional Genomics Core for assistance with sequencing. This work was supported by the National Institutes of Health (NIH) Director's Pioneer Award: DP1CA238296 to C.C. and the National Cancer Institute (NCI) Cancer Target Discovery and Development Center (U01CA217851) to C.K. and C.C. K.K. was supported in part by a Swedish Research Council (Ventenskapsradet) International postdoc grant (2018-00454).

Author contributions
Conceptualization: CC; Methodology: KK, MP, HX, EK, KrK, AS, KL, AM, CK, CC; Investigation: KK, MP, HX, EK, KrK, AK, ZM, YHL, BL, CB, CJS; Visualization: KK, MP, HX, EK, KEH, AK; Funding acquisition: CK, CC; Project administration: CC; Supervision: CK, CC; Writing – original draft: KK, MP, CC. Writing – review & editing: all authors

Competing interests
CK is a founder and stockholder for Surrozen Inc, Mozart Therapeutics and NextVivo Inc. C.C. is an advisor and stockholder in Grail, Ravel, DeepCell and an advisor to Genentech, Bristol Myers Squibb, 3T Biosciences and NanoString. All other authors declare no competing interests.

Data availability
Metadata, cellranger outputs will be available via Zenodo:
https://doi.org/10.5281/zenodo.6401895
WGS and scRNA-seq data will be available via dbGAP

Code availability
The computational methods, procedures and analyses summarized above are implemented in custom R, python and bash scripts will be available via Github: https://github.com/cancersysbio/gastric_organoid_evolution

Graphical abstract

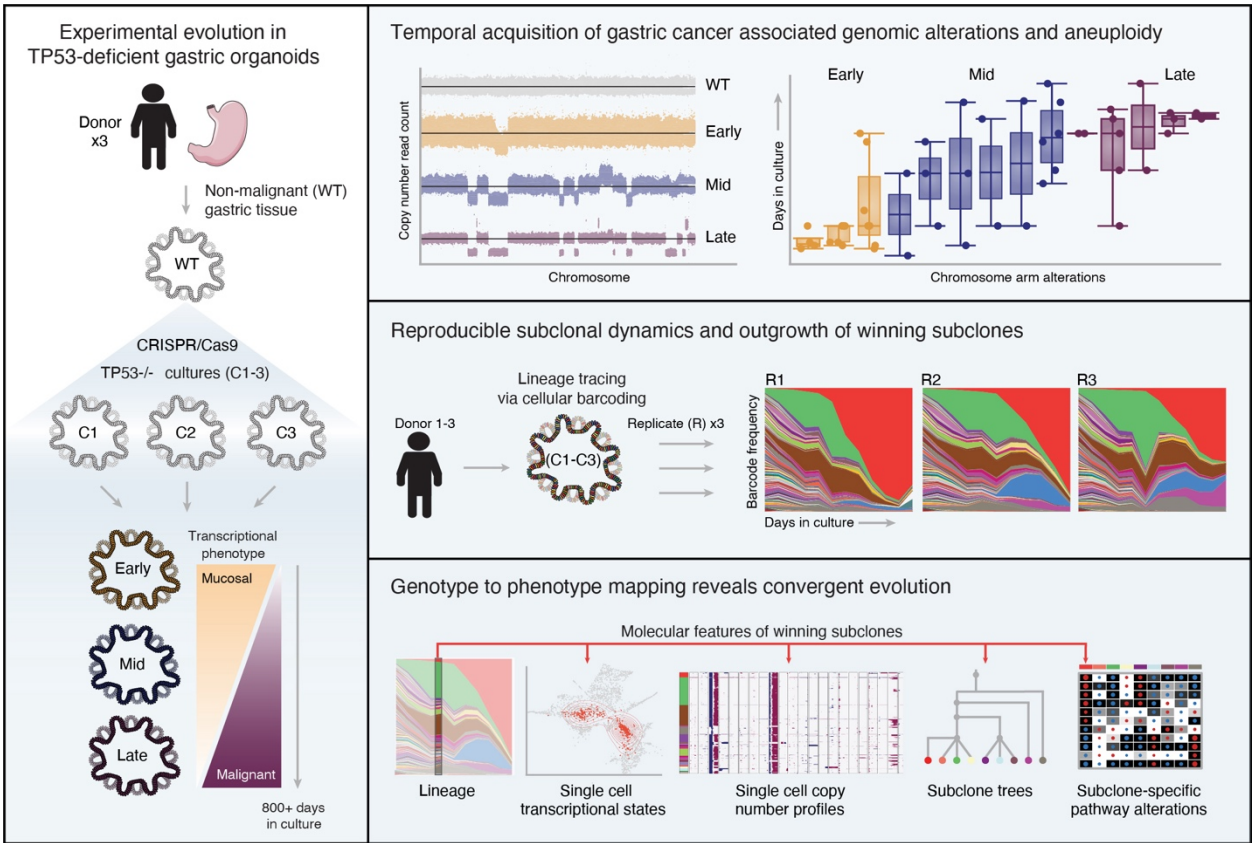

Figure 1

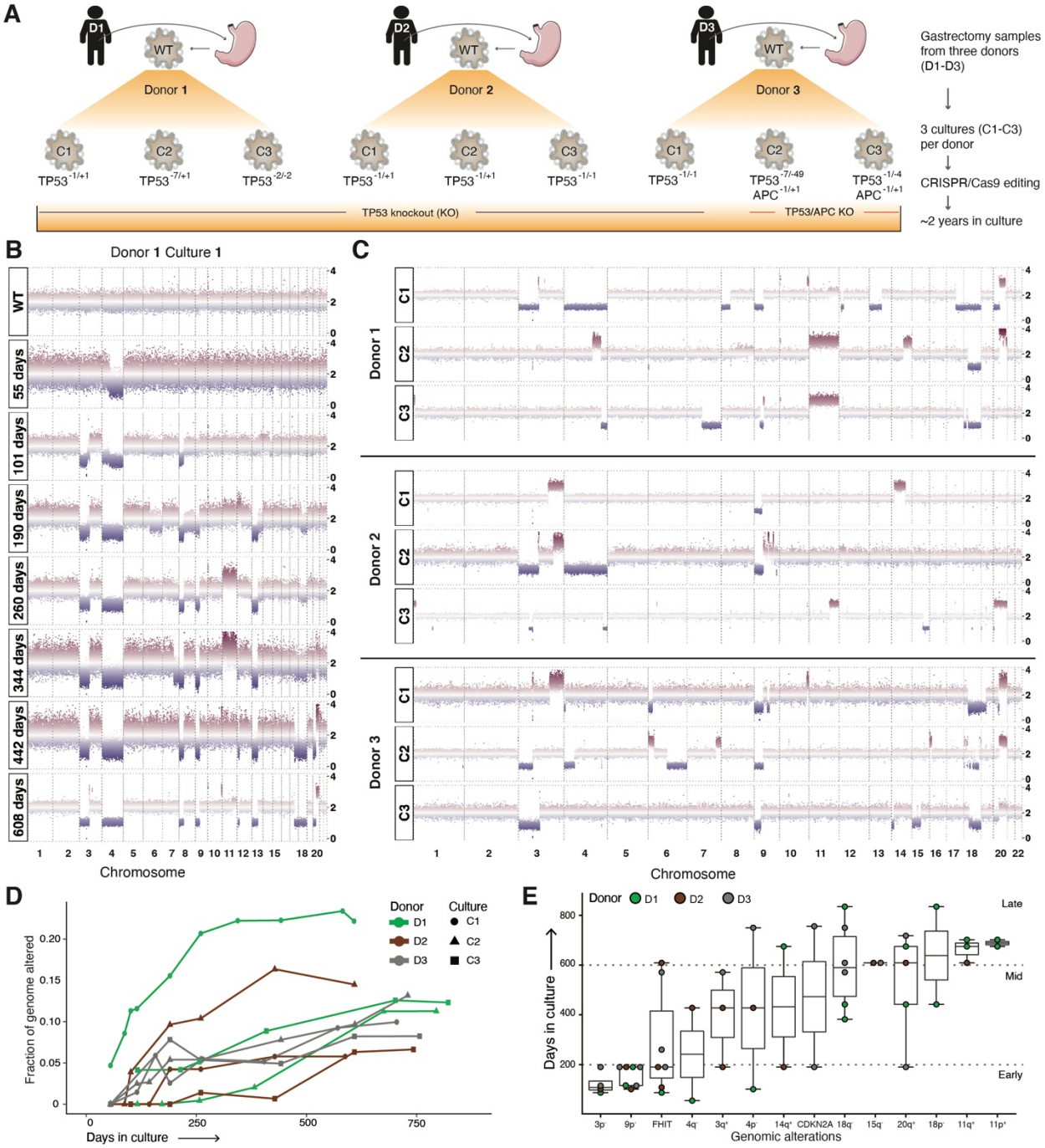
A
D1
D2
D3
WT
Donor 1
Donor 2
Donor 3
C1
C2
C3
TP53 knockout (KO)
TP53/APC KO
Gastrectomy samples from three donors (D1-D3)
3 cultures (C1-C3) per donor
CRISPR/Cas9 editing
~2 years in culture
B
Donor 1 Culture 1
WT
55 days
101 days
190 days
260 days
344 days
442 days
608 days
Chromosome
C
Donor 1
Donor 2
Donor 3
Chromosome
D
Fraction of genome altered
Days in culture
Donor
Culture
E
Donor
Late
Mid
Early
Days in culture
Genomic alterations

Fig. 1. *TP53* knock-out in gastric organoids induces aneuploidy and gastric cancer associated copy number alterations along a defined temporal order. (**A**) Schematic overview of organoid generation. Patient-derived organoids were established from gastrectomy tissue from three donors (D1 - D3). CRISPR/Cas9 editing of *TP53* and/or *APC* was performed independently, resulting in multiple clonally derived cultures with distinct frameshift mutations. (**B**) Genome-wide copy number aberration (CNA) profiles in Donor 1, Culture 1 (D1C1) assessed at multiple time points through shallow whole genome sequencing (sWGS). Normalized read counts across 50kb windows of the genome are shown for each timepoint. (**C**) Copy number profiles for the 9 organoid cultures sampled between days 588-835. (**D**) Fraction genome altered (FGA) over time for each culture. (**E**) Time of appearance (days in culture) of persistent arm-level CNAs (or alterations in *FHIT* and *CDKN2A*) in *TP53/APC*-deficient organoids (alterations that become extinct are not considered). The prevalence of these alterations in gastric cancer is summarized in Fig. S4B-C.

Figure 2

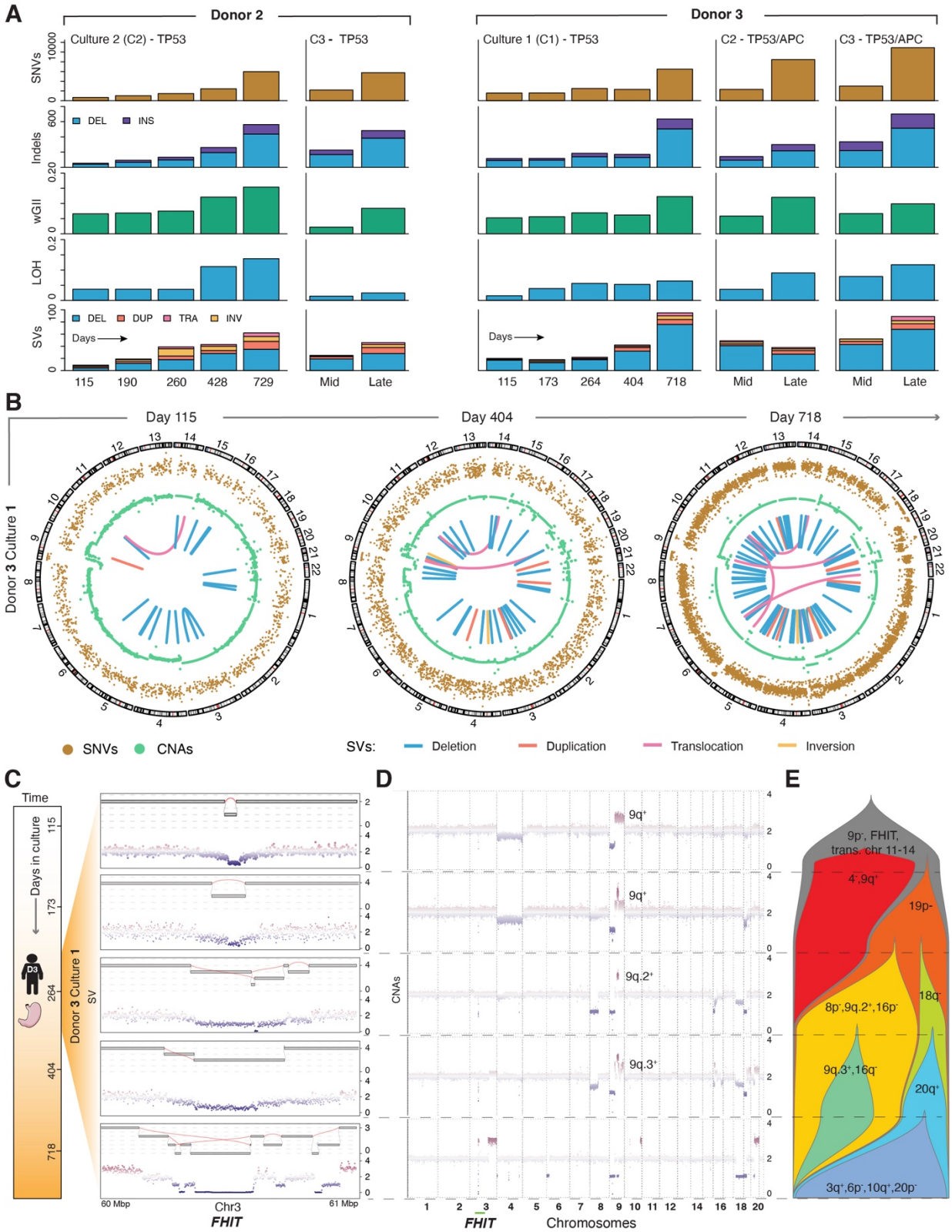
A
Donor 2
Culture 2 (C2) - TP53
C3 - TP53
Donor 3
Culture 1 (C1) - TP53
C2 - TP53/APC
C3 - TP53/APC
SNVs
Indels
wGII
LOH
SVs
DEL
INS
DUP
TRA
INV
Days
115
190
260
428
729
Mid
Late
173
264
404
718
B
Day 115
Day 404
Day 718
Donor 3 Culture 1
SNVs
CNAs
SVs:
Deletion
Duplication
Translocation
Inversion
C
Time
Days in culture
Donor 3 Culture 1
SV
60 Mbp
61 Mbp
Chr3
FHIT
D
CNAs
9q+
9q.2+
9q.3+
FHIT
Chromosomes
E
9p-, FHIT, trans. chr 11-14
4-,9q+
19p-
8p-,9q.2+,16p-
18q-
9q.3+,16q-
20q+
3q+,6p-,10q+,20p-

Fig. 2. *TP53* deficiency elicits subclonal copy number evolution and complex structural variation. (**A**) Burden of somatic genomic alterations in *TP53, TP53/APC* deficient gastric organoids (relative to paired WT) over time as assessed through longitudinal WGS of individual cultures at the specified time point (*Mid* - d296, *Late* d743-756). The classes of alterations and summaries assessed include: single nucleotide variants (SNV), insertions and deletions, weighted-genome instability index (wGII), fraction of genome with loss of heterozygosity (LOH), structural variants (SV) (transversions, inversions, duplications, and translocations). (**B**) Circos plots for D3C1 illustrates increasing genomic instability and complexity with time. Classes of alterations shown include: SNVs (adjusted variant allele frequencies), copy number alterations (logR) and SV consensus calls (see Methods). (**C**) Evolution of rigma-like SVs at the *FHIT* fragile site on chr3p. Zoomed in view of a 1Mb region in the *FHIT* locus. Reconstructed SVs are shown on the upper panel, with corresponding copy number profiles on the lower panel. (**D**) Longitudinal copy number aberration (CNA) profile for D3C1, for the corresponding time points in panel (C). (**E**) Fishplot schematic for D3C1 reveals subclonal copy number evolution, clonal interference and extinction. Subclone frequencies (x-axis) are determined based on the CNAs visualized in 2D (Methods).

Figure 3

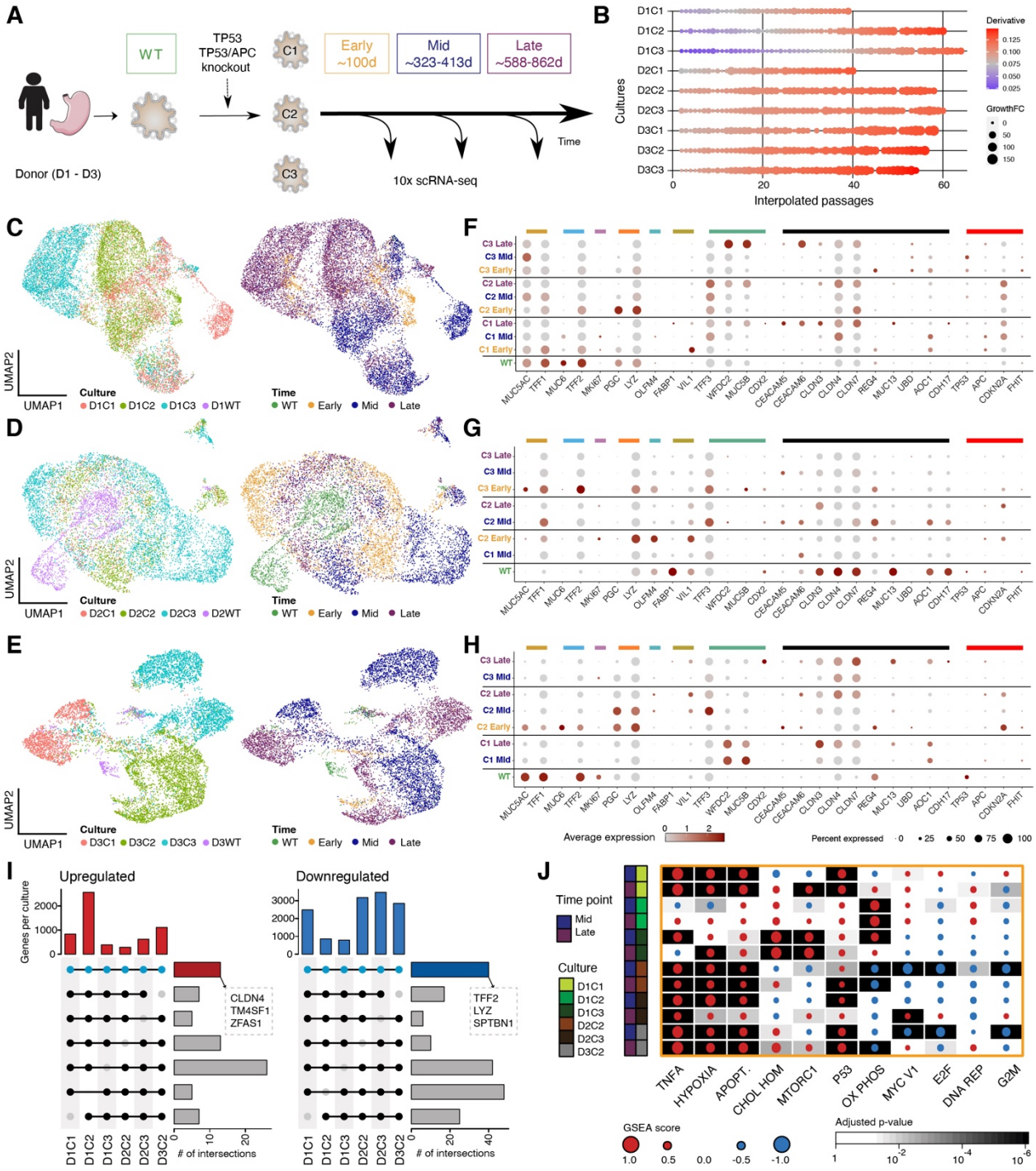
A
WT
TP53
TP53/APC
knockout
C1
C2
C3
Early
~100d
Mid
~323-413d
Late
~588-862d
Time
Donor (D1 - D3)
10x scRNA-seq
B
Cultures
Interpolated passages
Derivative
GrowthFC
C
D
E
UMAP1
UMAP2
Culture
D1C1 D1C2 D1C3 D1WT
D2C1 D2C2 D2C3 D2WT
D3C1 D3C2 D3C3 D3WT
Time
WT Early Mid Late
F
G
H
C3 Late
C3 Mid
C3 Early
C2 Late
C2 Mid
C2 Early
C1 Late
C1 Mid
C1 Early
WT
MUC5AC TFF1 MUC6 TFF2 MKI67 PGC LYZ OLFM4 FABP1 VIL1 TFF3 WFDC2 MUC5B CDX2 CEACAM5 CEACAM6 CLDN3 CLDN4 CLDN7 REG4 MUC13 UBD AOC1 CDH17 TP53 APC CDKN2A FHIT
Average expression
Percent expressed
I
Upregulated
Downregulated
Genes per culture
CLDN4
TM4SF1
ZFAS1
TFF2
LYZ
SPTBN1
# of intersections
D1C1 D1C2 D1C3 D2C2 D2C3 D3C2
J
Time point
Mid
Late
Culture
D1C1
D1C2
D1C3
D2C2
D2C3
D3C2
TNFA
HYPOXIA
APOPT.
CHOL HOM
MTORC1
P53
OX PHOS
MYC V1
E2F
DNA REP
G2M
GSEA score
Adjusted p-value

Fig. 3. Longitudinal single cell profiling reveals transcriptional deregulation following *TP53* loss. (**A**) Experimental overview of longitudinal single-cell RNA (scRNA-seq) profiling of gastric organoid cultures. Wild-type and three replicate *TP53, TP53/APC* KO gastric organoid cultures were sampled at multiple time points (*Early*, ~100 days, orange; *Mid*, ~320 days, blue; *Late*, ~770 days, purple) and subject to scRNA-seq. (**B**) Dotplot depicting estimated growth curve derivatives and growth fold-change (FC) from previous time point for each culture over time (interpolated passage number). (**C**, **D**, **E**) UMAP visualizations colored according to culture (left) and time point (right) for Donors 1, 2 and 3 depicting 13,984, 9,031 and 8,591 cells, respectively. (**F**, **G**, **H**) Dotplot depicting the expression of selected marker genes for individual cultures and time points shown in (C), (D) and (E). Colored bars highlight marker genes associated with normal gastric and intestinal cell types, genes up-regulated in the gene expression profiling interactive analysis (GEPIA) of gastric cancers, and other genes of functional relevance. Pit mucosal cells: *MUC5AC*, *TFF1* – dark yellow; Gland mucosal cells: *MUC6*, *TFF2* – light blue; Proliferative cells: *MKI67* – purple; Neck-like cells: *PGC*, *LYZ* – orange; Mucosal stem cells: *OLFM4* – turquoise; Enterocytes: *FABP1*, *VIL1* - olive; Goblet cells: *TFF3*, *WFDC2*, *MUC5B*, *CDX2* – green; GEPIA top 12 genes: *CEACAM5*, *CEACAM6*, *CLDN3*, *CLDN4*, *CLDN7*, *REG4*, *MUC3A*, *MUC13*, *PI3*, *UBD*, *AOC1*, *CDH17* - black; Functional importance: *TP53*, *APC*, *CDKN2A*, *FHIT* – red. (**I**) Upset plot representing shared differentially up- (left) and down-regulated genes (right) across donors and cultures. (**J**) Gene Set Enrichment Analysis (GSEA) heatmap for MSigDB Hallmark gene sets showing the most significantly altered pathways for each culture (Kolmogorov-Smirnov statistic, Benjamini-Hochberg adjusted). GSEA score is indicated (dot size) and colored according to the directionality of expression profiles (up, red; down, blue).

Figure 4

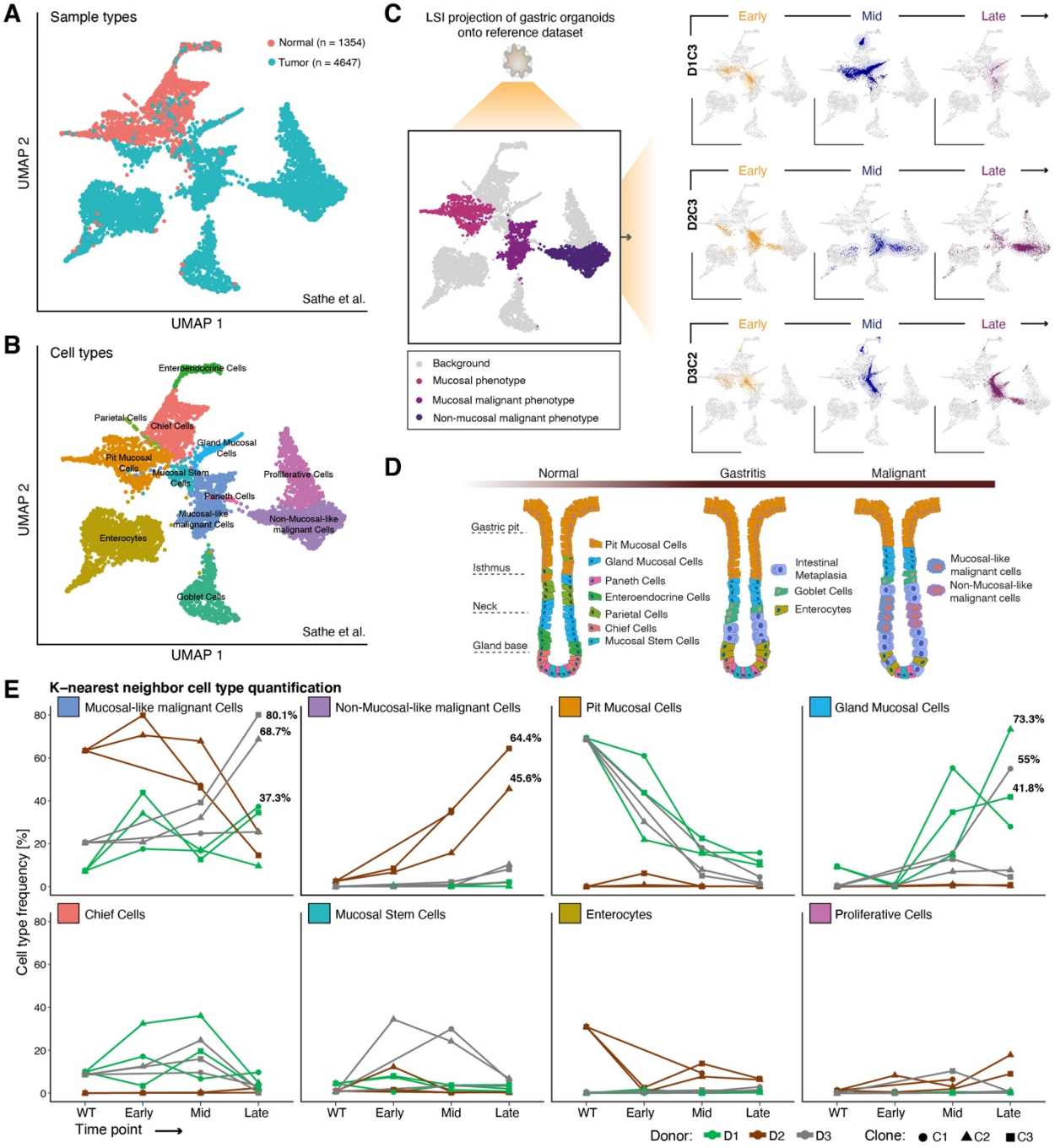
A
Sample types
Normal (n = 1354)
Tumor (n = 4647)
UMAP 2
UMAP 1
Sathe et al.
B
Cell types
Enteroendocrine Cells
Parietal Cells
Chief Cells
Gland Mucosal Cells
Pit Mucosal Cells
Mucosal Stem Cells
Proliferative Cells
Paneth Cells
Mucosal-like malignant Cells
Non-Mucosal-like malignant Cells
Enterocytes
Goblet Cells
C
LSI projection of gastric organoids onto reference dataset
Background
Mucosal phenotype
Mucosal malignant phenotype
Non-mucosal malignant phenotype
Early
Mid
Late
D1C3
D2C3
D3C2
D
Normal
Gastritis
Malignant
Gastric pit
Isthmus
Neck
Gland base
Pit Mucosal Cells
Gland Mucosal Cells
Paneth Cells
Enteroendocrine Cells
Parietal Cells
Chief Cells
Mucosal Stem Cells
Intestinal Metaplasia
Goblet Cells
Enterocytes
Mucosal-like malignant cells
Non-Mucosal-like malignant cells
E
K–nearest neighbor cell type quantification
Mucosal-like malignant Cells
Non-Mucosal-like malignant Cells
Pit Mucosal Cells
Gland Mucosal Cells
Chief Cells
Mucosal Stem Cells
Enterocytes
Proliferative Cells
80.1%
68.7%
37.3%
64.4%
45.6%
73.3%
55%
41.8%
Cell type frequency [%]
WT
Early
Mid
Late
Time point
Donor: D1 D2 D3
Clone: C1 C2 C3

Fig. 4. Unsupervised assessment reveals evolution towards malignant transcriptional states. (**A**, **B**) UMAP embedding of 6,080 epithelial cells from the Sathe et al. gastric tumor/normal scRNA-seq dataset, colored according to histology (A) and assigned cell type (B). Detected cell types included pit mucosal cells (PMCs), gland mucosal cells (GMCs), chief cells, parietal cells, enterocytes, enteroendocrine cells, goblet cells and proliferative cells, as well as two types of malignant cells (mucosal-like and non-mucosal-like). (**C**) Latent semantic indexing (LSI) projection (Methods) of *TP53*, *TP53*/*APC* KO gastric organoids sampled at *Early* (orange), *Mid* (blue) and *Late* (purple) timepoints onto the Sathe et al. reference dataset (Methods). The reference dataset (left) is colored by cellular phenotypes of interest, providing orientation for the LSI projection of the three gastric organoid cultures at the specified timepoints (right). The density of projected cells is highlighted using a 2D density distribution. (**D**) Schematic representation of shifts in cell populations proposed to accompany the transition from normal tissue to gastritis which can lead to intestinal metaplasia and ultimately malignancy, adapted from (*43*). (**E**) Projected cell type frequencies based on the 25 nearest neighbors (NNs) in WT and *TP53*, *TP53*/*APC* KO gastric organoids over time.

Figure 5

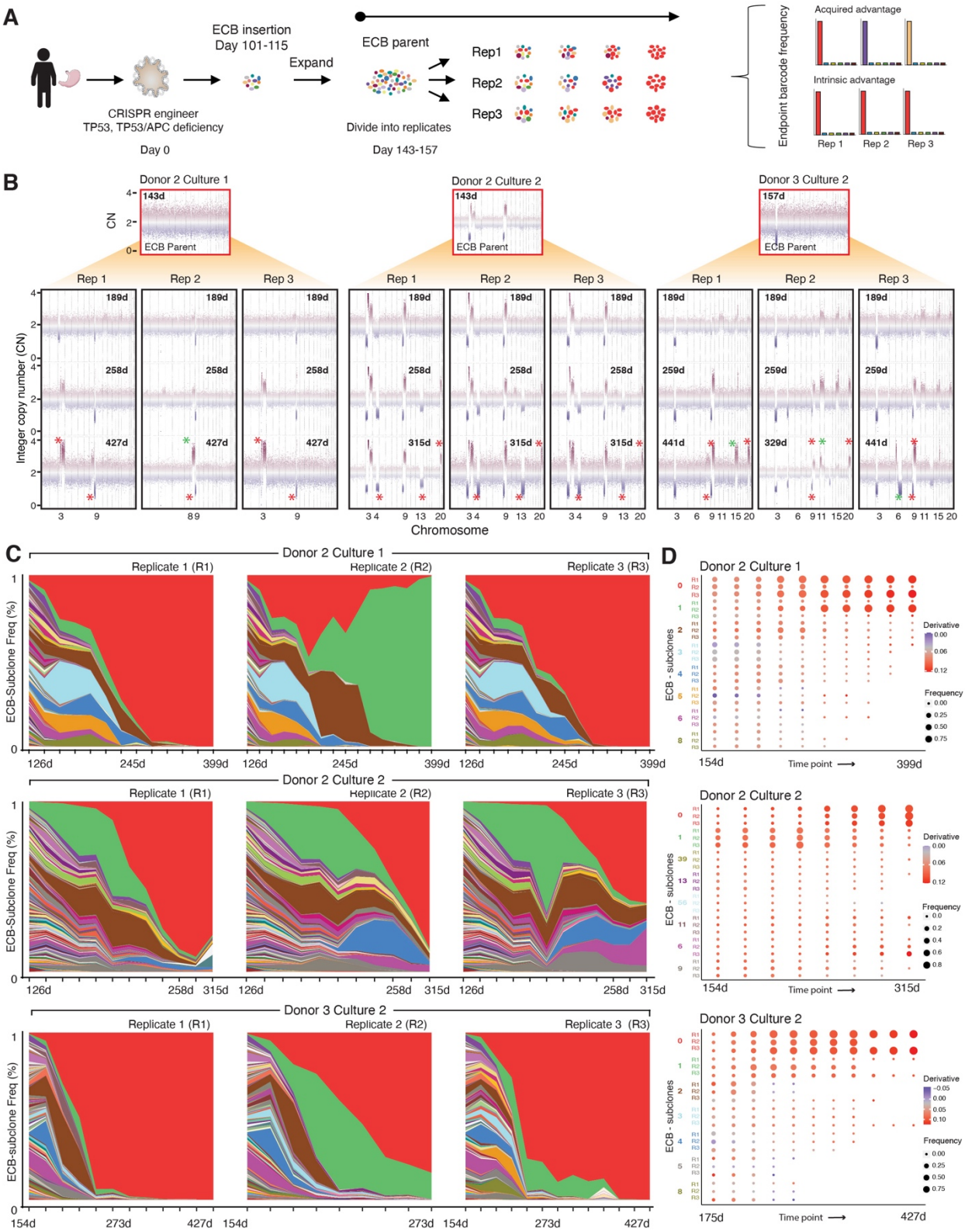
A
ECB insertion
Day 101-115
Expand
ECB parent
Rep1
Rep2
Rep3
CRISPR engineer
TP53, TP53/APC deficiency
Day 0
Divide into replicates
Day 143-157
Endpoint barcode frequency
Acquired advantage
Intrinsic advantage
Rep 1
Rep 2
Rep 3
B
Donor 2 Culture 1
Donor 2 Culture 2
Donor 3 Culture 2
CN
ECB Parent
Integer copy number (CN)
Chromosome
C
Replicate 1 (R1)
Replicate 2 (R2)
Replicate 3 (R3)
ECB-Subclone Freq (%)
D
ECB - subclones
Time point
Derivative
Frequency

Fig. 5. Lineage tracing reveals subclonal dynamics and deterministic outgrowth. (**A**) Experimental overview of prospective lineage tracing studies in *TP53*-/- gastric organoids using expressed cellular barcodes (ECBs). Lentiviral transduction of a high-complexity ECB library into *TP53* or *TP53*/*APC* KO gastric organoid cultures 'tags' each cell with a unique barcode (Methods). The ECB parental population was split into replicates, and individual cultures evolved in parallel and subject to longitudinal barcode sequencing, revealing subclonal dynamics and assessment of intrinsic or acquired fitness advantages amongst replicate cultures. (**B**) Copy number aberration (CNA) profiles were assessed by shallow WGS (sWGS) prior to introduction of the ECB in the parental line and across replicate ECB cultures at multiple time points over ~two years. Red stars denote CNAs present in at least two replicates but not in the parental population; green stars denote CNAs unique to one replicate. Only chromosomes that harbor newly arising CNAs (not present in the parental population) are numbered for simplicity. (**C**) Mueller plots depict ECB frequencies (assessed by barcode sequencing) over time where each color represents a distinct subclone in each replicate. (**D**) Dotplots indicate ECB subclone frequency (size) and estimated growth curve derivative per subclone (color).

Figure 6

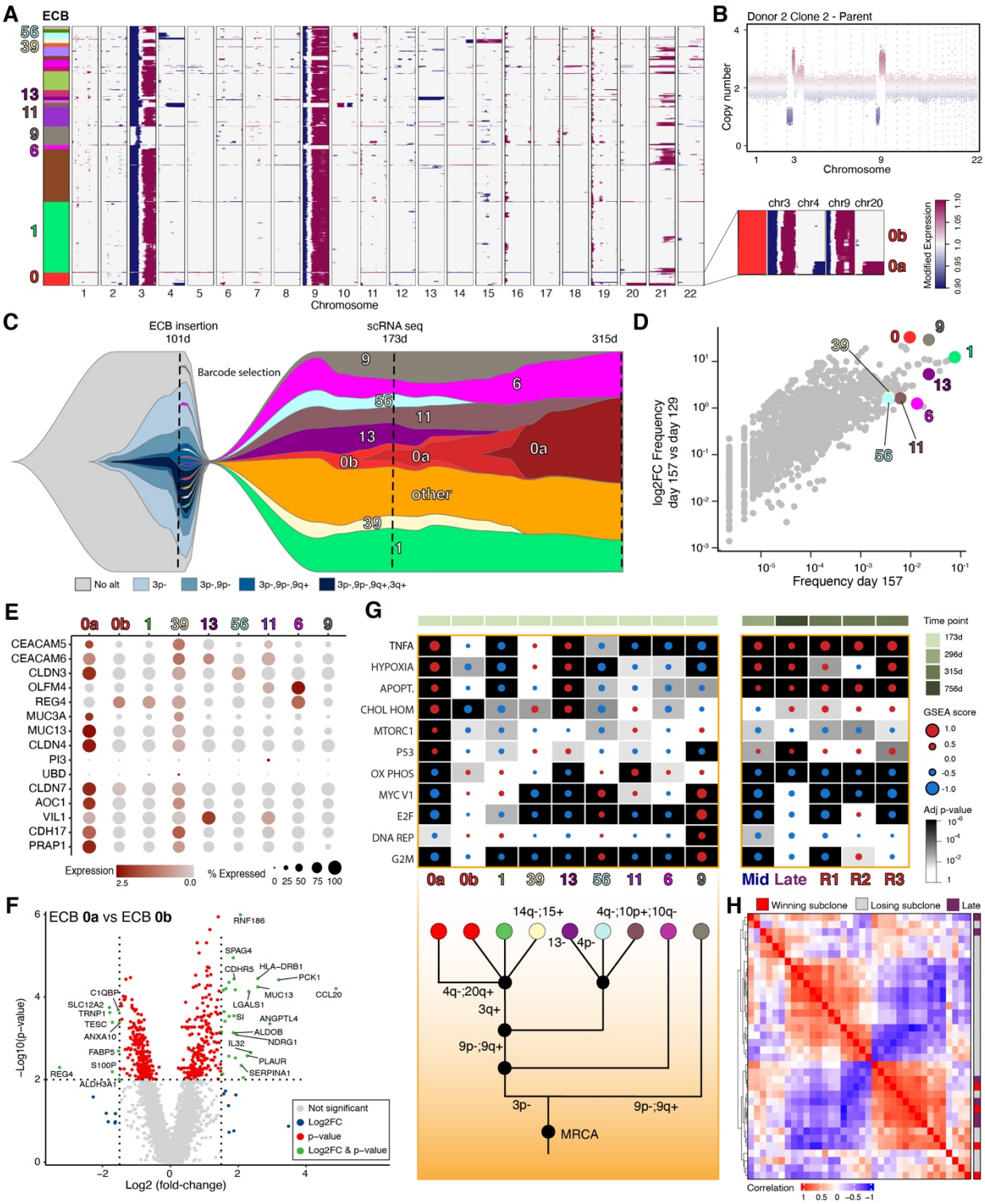
A
ECB
Chromosome
B
Donor 2 Clone 2 - Parent
Copy number
Chromosome
chr3 chr4 chr9 chr20
Modified Expression
C
ECB insertion
101d
scRNA seq
173d
315d
Barcode selection
other
No alt
3p-
3p-,9p-
3p-,9p-,9q+
3p-,9p-,9q+,3q+
D
log2FC Frequency
day 157 vs day 129
Frequency day 157
E
CEACAM5
CEACAM6
CLDN3
OLFM4
REG4
MUC3A
MUC13
CLDN4
PI3
UBD
CLDN7
AOC1
VIL1
CDH17
PRAP1
Expression
% Expressed
F
ECB 0a vs ECB 0b
Log2 (fold-change)
−Log10(p-value)
Not significant
Log2FC
p-value
Log2FC & p-value
G
TNFA
HYPOXIA
APOPT.
CHOL HOM
MTORC1
P53
OX PHOS
MYC V1
E2F
DNA REP
G2M
Time point
GSEA score
Adj p-value
Mid Late R1 R2 R3
14q-;15+
4q-;10p+;10q-
13-
4p-
4q-;20q+
3q+
9p-;9q+
3p-
MRCA
H
Winning subclone
Losing subclone
Late
Correlation

Fig. 6. Genotype to phenotype mapping defines molecular determinants of winning subclones. (**A**) Inferred copy number aberration heatmap from scRNA-seq data for Donor 2, Culture 2, Replicate 2 (D2C2R2) at day 173, where each row is a cell. The color bar at the left indicates the expressed cellular barcode (ECB) each cell maps to. Numbered barcodes denote those selected for further investigation. The inset plot shows that a subpopulation within the winning red subclone (ECB-0) acquires additional CNAs, including deletion of chr4q and gain of 20q and is termed "0a", while the ECB-0 parent subclone is termed "0b" (**B**) Copy number aberration (CNA) profile for the D2C2 bulk parental population based on sWGS for comparison (also shown in Fig 5B). (**C**) Fishplot schematic illustrating the link between lineage (ECBs) and CNA based subclones. To facilitate visualization, subclones of interest (denoted in A) are shown, while the remainder grouped as "*other*". Actual subclone frequencies are depicted in Fig 5C. the winning red subclone (ECB-0a) which derives from ECB-0b acquires nested CNAs in which all cells harboring the second CNA also harbor the first (depicted by darker shading). (**D**) Scatterplot comparing subclone frequency at day 157 and the log2 fold change between days 129 and 157. All subclones are shown, those of interest are highlighted as in panel A. (**E**) Dotplot showing the expression of top differentially expressed genes (DEGs) based on gene expression profiling interactive analysis (GEPIA) of gastric cancers (GC). (**F**) Volcano plot illustrating DEGs from the comparison of the winning subclone 0a and its parent subclone 0b. Vertical and horizontal lines correspond to absolute log2FC values of 1.5 and $p < 0.01$, respectively. (**G**) Gene Set Enrichment Analysis (GSEA) heatmap from MSigDB Hallmark gene sets showing the most significantly altered pathways for specific subclone at day 173 (left; Kolmogorov-Smirnov statistic, Benjamini-Hochberg adjusted). GSEA score is colored according to the type of differential expression (up = red; down = blue). A manually reconstructed copy number phylogeny is shown below. The heatmap on the right shows a comparison between subclone 0a and all other cells at a later time point (day 315) for all three replicates (R1-R3) as well as the comparison of D2C2 *Mid* (day 323) and *Late* (day 756) time points relative to *Early* (day 101), as in Fig. 3J. (**H**) Pairwise spearman correlation between samples based on the GSEA score for the top 10 most altered pathways for *Late* relative to *Early* timepoints and for winning or losing subclones where DEGs within each subclone were compared to all other cells at that timepoint. The most significantly altered pathways are shown in Fig. S22 and an extended version of this heatmap with individual cultures labeled is shown in Fig. S23.