









TRIM28-dependent developmental heterogeneity determines cancer susceptibility through distinct epigenetic states

Received: 29 August 2023

Accepted: 13 December 2024

Published online: 24 January 2025

 Check for updates

Ilaria Panzeri ^{1,2} , Luca Fagnocchi¹, Stefanos Apostle¹, Megan Tompkins³, Emily Wolfrum ⁴, Zachary Madaj⁴, Galen Hostetter⁵, Yanqing Liu¹, Kristen Schaefer ^{1,6}, Chih-Hsiang Yang^{1,2,7}, Alexis Bergsma^{1,8}, Anne Drougard¹, Erez Dror², PERMUTE^{*}, Darrell P. Chandler¹, Daniel Schramek ^{9,10}, Timothy J. Triche Jr. ^{1,11,12} & John Andrew Pospisilik ^{1,2} 

Mutations in cancer risk genes increase susceptibility, but not all carriers develop cancer. Indeed, while DNA mutations are necessary drivers of cancer, only a small subset of mutated cells go on to cause the disease. To date, the mechanisms underlying individual cancer susceptibility remain unclear. Here, we took advantage of a unique mouse model of intrinsic developmental heterogeneity (*Trim28^{+/D9}*) to investigate whether early-life epigenetic variation influences cancer susceptibility later in life. We found that heterozygosity of *Trim28* is sufficient to generate two distinct early-life epigenetic states associated with differing cancer susceptibility. These developmentally primed states exhibit differential methylation patterns at typically silenced heterochromatin, detectable as early as 10 days of age. The differentially methylated loci are enriched for genes with known oncogenic potential, frequently mutated in human cancers and correlated with poor prognosis. This study provides genetic evidence that intrinsic developmental heterogeneity can prime individual, lifelong cancer susceptibility.

Cancer is triggered by oncogenic DNA mutations¹. However, these mutations are also found at high rates in otherwise ‘healthy’ tissues, and not every mutation is oncogenic across all tissues^{2–5}. Thus, the oncogenic potential of DNA mutations is cell, tissue and temporal specific^{6,7}. The molecular basis of this context specificity comprises one of the biggest unanswered questions in cancer biology.

Pioneering studies over the last decades have implicated epigenetic regulation as a key mediator of tumorigenesis. Notable examples include demonstrations that cell type- and differentiation stage-specific differences in epigenetic control determine when and

where transformation occurs^{8–10}, and parallel to this study, transient developmental perturbation of Polycomb-dependent epigenetic silencing in *Drosophila* was shown to induce irreversible cell fate changes that support oncogenesis¹¹. Collectively, these data suggest that tumors can emerge as a result of developmental epigenetic (dys) regulation. One layer of epigenetic variation that is largely overlooked in human epidemiology is ‘intrinsic developmental heterogeneity’. This layer comprises epigenetic differences that arise during development in part through stochastic processes¹² that are distinct from the epigenetic changes triggered by external environmental exposures

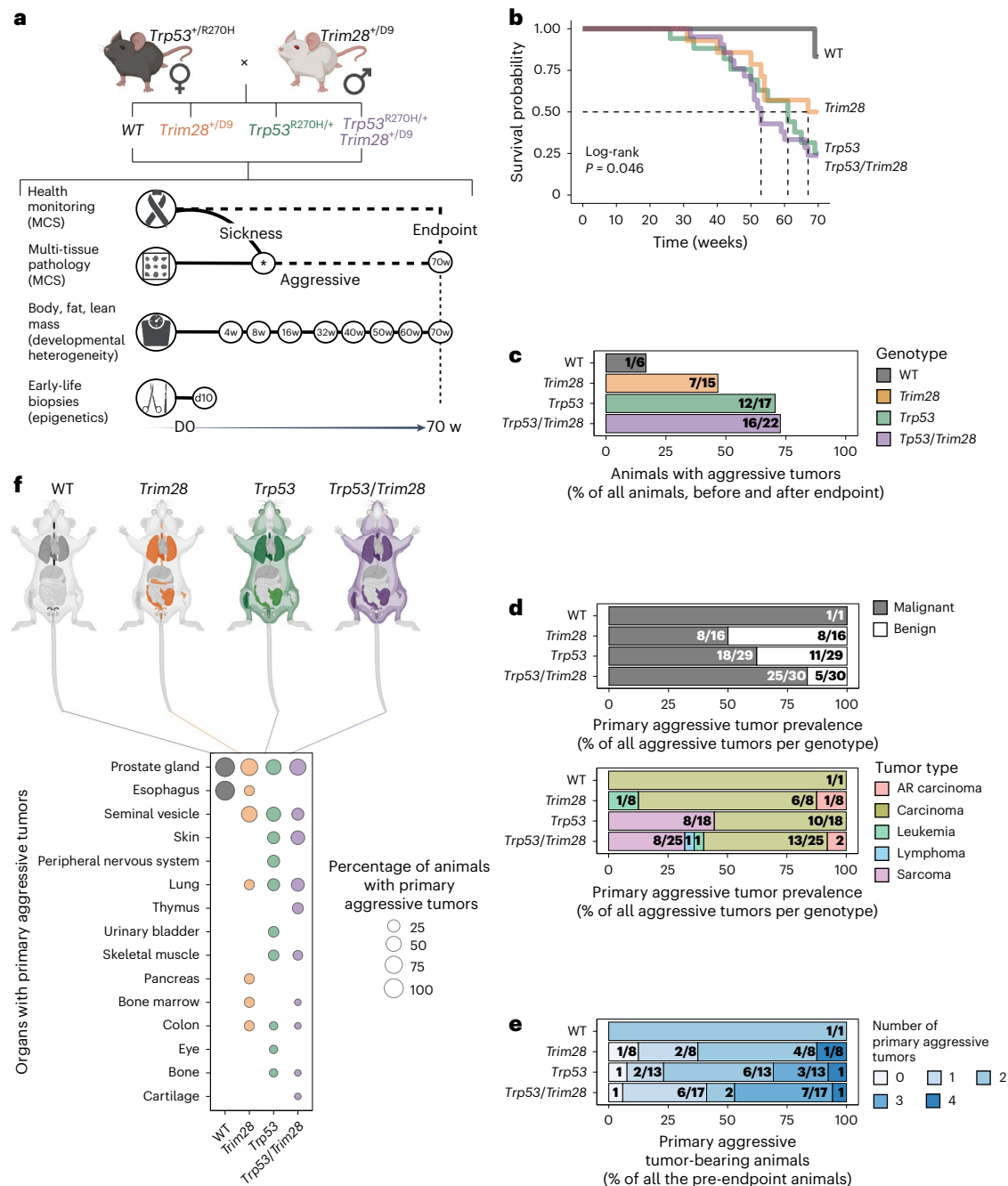


Fig. 1 | *Trim28*^{+D9} mice exhibit MCS. a, Schematic of the experimental plan. The *Trp53*^{R270H/+} MCS model was mated with the *Trim28*^{+D9} developmental heterogeneity model. F₁ genotypes were screened for health issues and tumor development. Tissues were collected before (aggressive) or at the endpoint of the study (70 weeks (w)). Histopathology determined the presence of tumors. Body, fat and lean mass was measured at multiple time points. Early-life biopsies were collected at day (D) 10 (before weaning). **b**, Kaplan–Meier survival probability by genotype. Log-rank test, *P* = 0.046. *n* = 60 male mice screened for cancer (6 wild type (WT), 15 *Trim28*^{+D9}, 17 *Trp53*^{R270H/+}, 22 *Trp53*^{R270H/+};*Trim28*^{+D9}). **c**, Prevalence of male mice with aggressive tumors, as percentage relative to total screened males in each genotype (before and after endpoint). Actual numbers are given in fractions. *n* = 60 males (6 WT, 15 *Trim28*^{+D9}, 17 *Trp53*^{R270H/+}, 22 *Trp53*^{R270H/+};*Trim28*^{+D9}). **d**, Top, prevalence of malignant (gray) or benign (white) aggressive tumors (before endpoint) for each genotype, as a percentage relative to total aggressive tumors by genotype. Actual numbers are given in fractions. *n* = 76 malignant and benign aggressive tumors (1 in WT, 16 in *Trim28*^{+D9}, 29 in *Trp53*^{R270H/+}, 30 in *Trp53*^{R270H/+};*Trim28*^{+D9}). Bottom, prevalence of distinct

malignant aggressive tumor types (before endpoint) in each genotype as a percentage relative to total malignant aggressive tumors (before endpoint). Actual numbers are given in fractions (if the denominator is not specified due to space constraints, consider it the same within each genotype). *n* = 52 malignant aggressive tumors (1 in WT, 8 in *Trim28*^{+D9}, 18 in *Trp53*^{R270H/+}, 25 in *Trp53*^{R270H/+};*Trim28*^{+D9}). AR, age-related. **e**, Fraction of mice with 0, 1 or multiple malignant aggressive tumors (before endpoint), relative to total pre-endpoint mice. Actual numbers are given in fractions (if the denominator is not specified due to space constraints, consider it the same within each genotype). *n* = 39 pre-endpoint mice (1 WT, 8 *Trim28*^{+D9}, 13 *Trp53*^{R270H/+}, 17 *Trp53*^{R270H/+};*Trim28*^{+D9}). **f**, Tissues affected by malignant aggressive tumors (before endpoint). Top, mouse anatomy plots; nontargeted tissues are in light gray, and targeted tissues are colored by genotype: WT (black), *Trim28*^{+D9} (orange), *Trp53*^{R270H/+} (green), *Trp53*^{R270H/+};*Trim28*^{+D9} (purple). Bottom, percentage of mice with organs targeted by malignant aggressive tumors (before endpoint) by genotype. *n* = 39 pre-endpoint mice (1 WT, 8 *Trim28*^{+D9}, 13 *Trp53*^{R270H/+}, 17 *Trp53*^{R270H/+};*Trim28*^{+D9}). Panels a, f (top) created with BioRender.com.

or from in utero insults that can increase cancer risk (for example, estrogens, alcohol and overnutrition or undernutrition)^{13–16}. This intrinsic developmental heterogeneity occurs at rates several orders of magnitude higher than those of mutations¹⁷ and can drive phenotypic discordance, for instance, between monozygotic twins or isogenic mice^{17–19}. While an impressive theoretic framework has been developed for how intrinsic developmental heterogeneity might impact cancer²⁰, to our knowledge, the notion has never been tested experimentally.

Because tumor initiation involves some degree of randomness, testing the relationship between developmental heterogeneity and cancer susceptibility requires measurement of the distribution of observed outcomes comparing distinct intrinsic epigenetic states: it requires an isogenic model, raised in tightly controlled environments but bearing more than one reproducible intrinsic epigenetic state^{21,22}. TRIM28 (also known as KAP1) is an epigenetic regulator that has an important role in heterochromatic gene silencing^{23,24}. TRIM28 loss-of-function models have implicated the protein in cancer in a complex and tissue-specific manner²⁵. TRIM28, however, is also a master regulator of organism-level developmental heterogeneity²⁶. Our prior work showed that genetically and environmentally identical *Trim28*^{+/-} *D9* haploinsufficient mice emerge into adulthood in one of two distinct phenotypes (or developmental morphs) characterized by differences in body mass composition²⁷. The *Trim28*^{+/-} *D9* mouse thus provides a sensitized model to detect the long-term phenotypic consequences of two distinct intrinsic developmental heterogeneity states. Here, we leverage this unique model to test whether intrinsic developmental heterogeneity regulates cancer susceptibility. We show that the two *Trim28*^{+/-} *D9* developmental morphs develop distinct timing, severity and types of cancer. We identify a signature of DNA hypomethylated genes, installed before weaning, that stratify mice for cancer susceptibility and outcome. These same genes are frequently mutated in human cancers, and their dysregulation correlates with poor prognosis, suggesting that, if conserved, this mode of action has the potential to impact a broad portion of the population.

Results

Trim28^{+/-} *D9* mice exhibit multi-cancer syndrome

To test whether intrinsic developmental heterogeneity influences cancer, we crossed B6J. *Trp53*^{+/-} *R270H* mice with FVB. *Trim28*^{+/-} *D9* mice (Fig. 1a). For consistency of parent-of-origin effects, we crossed *Trp53*^{+/-} *R270H* females with *Trim28*^{+/-} *D9* males. The *Trp53*^{+/-} *R270H* mouse is a multi-cancer syndrome (MCS) model¹⁸, while the *Trim28*^{+/-} *D9* mouse is sensitized to exhibit reproducible bistable developmental heterogeneity^{26,27}. Both lines were highly backcrossed, and cohorts yielded isogenic offspring at expected Mendelian ratios (Extended Data Fig. 1a): wild type, *Trim28*^{+/-} *D9* single heterozygotes (*Trim28*), *Trp53*^{+/-} *R270H* single heterozygotes (*Trp53*) and *Trp53*^{+/-} *R270H*; *Trim28*^{+/-} *D9* compound heterozygotes (*Trp53/Trim28*). Parental ID, litter size and housing density were all carefully recorded to reduce confounders. We tracked mice from birth to endpoint (70 weeks

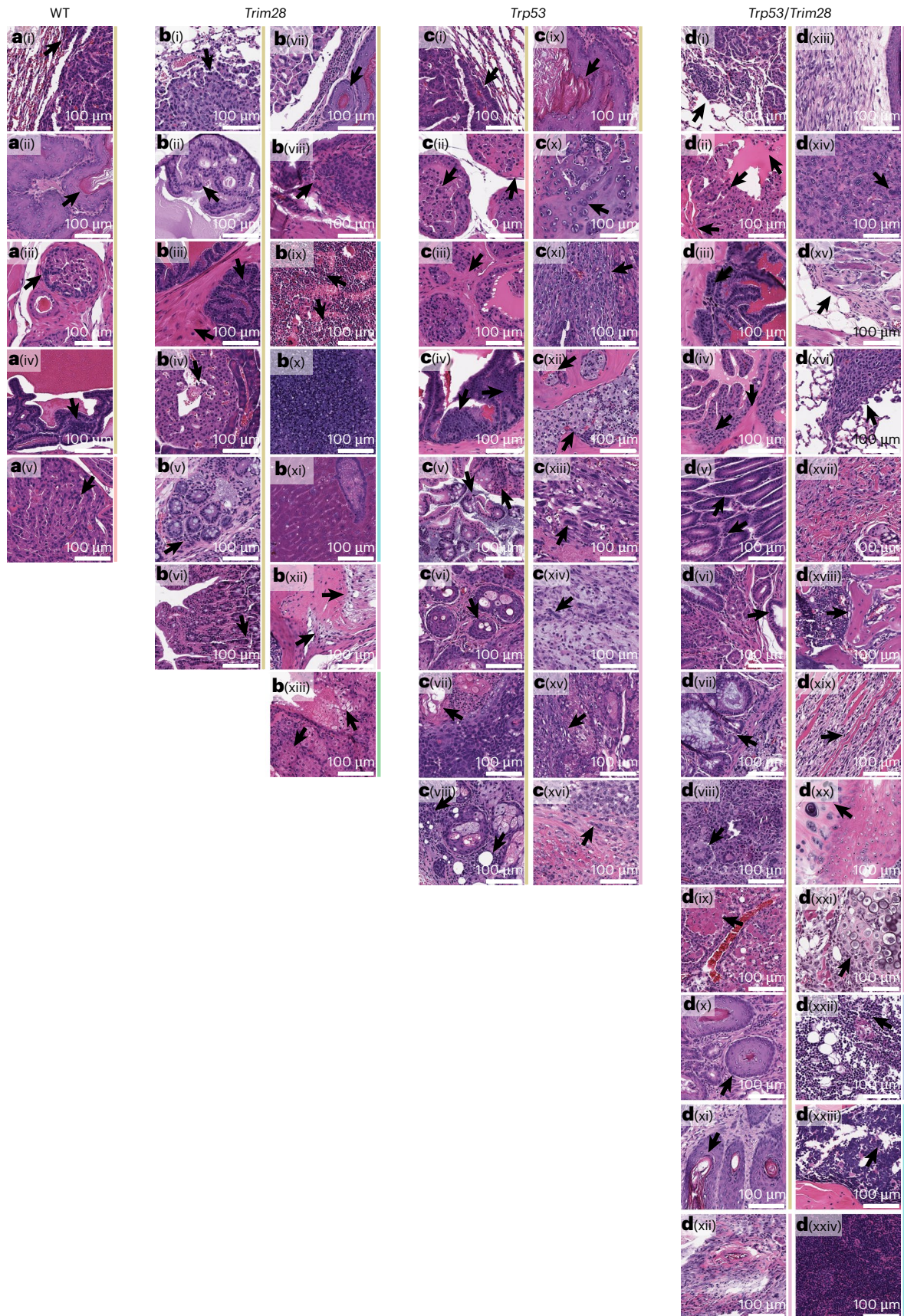
of age), monitoring each individual for sickness two to three times per week, with periodic measures of morphological, growth and metabolic characteristics (Fig. 1a). Early-life ear biopsies were obtained at 10 days of age for epigenomic profiling. Upon euthanasia, all mice underwent a systematic 21-organ dissection protocol in which tissues were isolated, processed for histology and scored by a pathologist. Cancer events were divided into aggressive (that is, mice requiring euthanasia before 70 weeks) and endpoint (that is, mice reaching the 70-week endpoint without evidence of sickness) events. The final dataset comprised 137 male mice with 79 malignant and 42 benign primary tumors and 133 female mice with 92 malignant and 36 benign primary tumors.

As expected, *Trp53*^{+/-} *R270H* male siblings exhibited high-penetrance MCS, with 76% of *Trp53*^{+/-} *R270H* mice succumbing to aggressive tumors before the 70-week endpoint (Fig. 1b,c). We found 24 primary malignant tumors in *Trp53*^{+/-} *R270H* males (mainly carcinomas and sarcomas), 18 of which developed before the 70-week endpoint (Fig. 1c). The tumors were found distributed throughout the body (Fig. 1d–f and Extended Data Fig. 1c–f). Surprisingly, *Trim28*^{+/-} *D9* male heterozygotes also showed reduced survival probability, similar to *Trp53*^{+/-} *R270H* male mice (Fig. 1b; mean survival of 59.9 and 56.5 weeks, respectively). Health monitoring and histopathology revealed that the reduced survival of *Trim28*^{+/-} *D9* mice was also due to MCS. Intriguingly, *Trim28*^{+/-} *D9* male mice showed time to detection and tumor burden similar to those of *Trp53*^{+/-} *R270H* male mice (Fig. 1b–d, top and Fig. 1e). That said, the *Trim28*^{+/-} *D9*-triggered MCS showed several differences relative to that of *Trp53*^{+/-} *R270H*. First, *Trim28*^{+/-} *D9* tumors included rare germ cell tumors (Extended Data Fig. 1c,d,f, right). Second, *Trim28*^{+/-} *D9* male mice showed only a single sarcoma, whereas sarcomas were common in *Trp53*^{+/-} *R270H* male mice (Fig. 1d, bottom and Extended Data Fig. 1d). Overall, carcinomas dominated the landscape of malignant primary tumors across genotypes (representative histology, Fig. 2). *Trp53*^{+/-} *R270H*; *Trim28*^{+/-} *D9* compound heterozygotes showed the lowest survival probability of all genotypes (Fig. 1b), with target tissue and tumor type distributions consistent with the presence of both alleles (Fig. 1f,d). No evidence was detected of a substantial genetic interaction between the two alleles when considering overall survival or tissue distribution. Stratification revealed similar pathologies between aggressive and endpoint samples, suggesting a largely constant etiological process (Fig. 1b–e and Extended Data Fig. 1b–f). Age-associated carcinomas were detected and overrepresented in wild-type male mice, as expected (Extended Data Fig. 1d,f, right). The few sarcomas that were observed in *Trim28*^{+/-} *D9* male mice were found at the endpoint (Extended Data Fig. 1d). Of note, *Trim28*^{+/-} *D9* male mice exhibited a higher number of benign tumors than other genotypes before the endpoint of the study (Fig. 1d). However, these were significantly reduced by the end of the study (Extended Data Fig. 1b), suggesting that the development of malignant tumors may be temporarily delayed in *Trim28*^{+/-} *D9* compared to *Trp53*^{+/-} *R270H* male mice. Thus, *Trim28*^{+/-} *D9* triggers an MCS similar to that of *Trp53*^{+/-} *R270H*.

Fig. 2 | Representative histological examples of tumor types for all genotypes and time points.

Arrows indicate main features of each tumor type. Tumor categories are color coded on the right: age-related carcinoma (red), carcinoma (gold), leukemia (light green), lymphoma (light blue), sarcoma (pink), germ cell tumors (dark green). **a**, WT male mice. **a**(i), Bronchoalveolar carcinoma (BAC). **a**(ii), Gastro-esophageal junction SCC. **a**(iii), Prostatic carcinoma (PCa). **a**(iv), Seminal vesicle carcinoma. **a**(v), Age-related PCa. **b**, *Trim28*^{+/-} *D9* male mice. **b**(i), BAC. **b**(ii), PCa. **b**(iii), Seminal vesicle carcinoma. **b**(iv), Pancreatic carcinoma. **b**(v), Large colon ACA. **b**(vi), Small bowel carcinoma. **b**(vii), Gastro-esophageal junction SCC. **b**(viii), Bladder transitional cell carcinoma. **b**(ix), AML affecting the bone marrow. **b**(x), AML spreading to the spleen. **b**(xi), AML spreading to the liver. **b**(xii), Synovial sarcoma. **b**(xiii), Seminoma. **c**, *Trp53*^{+/-} *R270H* male mice. **c**(i), BAC. **c**(ii), Age-related PCa. **c**(iii), PCa. **c**(iv), Seminal vesicle carcinoma. **c**(v), Large colon ACA. **c**(vi), Skin SCC. **c**(vii), Preputial gland SCC. **c**(viii), Adenoid cystic carcinoma. **c**(ix), Gastro-esophageal junction SCC. **c**(x), Osteosarcoma.

c(xi), Fibrosarcoma. **c**(xii), Chondrosarcoma. **c**(xiii), Rhabdomyosarcoma. **c**(xiv), Malignant peripheral nerve sheath tumor. **c**(xv), Sarcoma affecting the preputial glands. **c**(xvi), Dermatofibrosarcoma protuberans. **d**, *Trp53*^{+/-} *R270H*; *Trim28*^{+/-} *D9* male mice. **d**(i), BAC. **d**(ii), PCa. **d**(iii), Seminal vesicle carcinoma. **d**(iv), Age-related PCa. **d**(v), Gastro-esophageal junction ACA. **d**(vi), Duodenal carcinoma. **d**(vii), Colorectal carcinoma. **d**(viii), Thymic carcinoma. **d**(ix), Hepatoid ACA of the lung. **d**(x), Gastro-esophageal junction SCC. **d**(xi), Skin SCC. **d**(xii), Dermatofibrosarcoma protuberans. **d**(xiii), Soft tissue sarcoma. **d**(xiv), Epithelioid fibrohistiocytic sarcoma. **d**(xv), Skin appendage tumor. **d**(xvi), Metastatic fibrosarcoma to the lungs. **d**(xvii), Fibrosarcoma. **d**(xviii), Osteosarcoma. **d**(xix), Rhabdomyosarcoma. **d**(xx), Chondrosarcoma in the bone. **d**(xxi), Chondrosarcoma spreading to soft tissue. **d**(xxii), T cell lymphoma. **d**(xxiii), Acute lymphoblastic leukemia affecting the bone marrow. **d**(xxiv), Acute lymphoblastic leukemia spreading to the spleen.



Developmental heterogeneity primes cancer outcomes

Consistent with our previous work²⁷, *Trim28*^{+/D9} males showed marked variation in body mass at 16 weeks of age (Fig. 3a), and these separated statistically (mclust) into two distinct developmental ‘morphs’ (reproducible phenotypic forms): *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light (Fig. 3b,c and Extended Data Fig. 2a,b). *Trp53*^{R270H/+}; *Trim28*^{+/D9} compound heterozygotes also showed high variation in body mass, indicating that the intrinsic developmental heterogeneity triggered by *Trim28*^{+/D9} is maintained in the presence *Trp53*^{R270H/+}; however, bimodality could not be statistically resolved (Fig. 3b,c and Extended Data Fig. 2a,b). Neither wild-type nor *Trp53*^{R270H/+} heterozygotes showed a significant variance phenotype (Fig. 3a,b) or bimodality (Fig. 3a–c and Extended Data Fig. 2a,b). We validated the annotation of morphs using Rmixmod and found 100% congruence with mclust (Extended Data Fig. 2c). Notably, these phenotypic distinctions between morphs were transient over the long term (Extended Data Fig. 2a). The developmental bifurcation in *Trim28*^{+/D9} mice was critical to our experimental design, because it enables statistical comparison of cancer outcomes between groups of isogenic mice with reproducible and distinct intrinsic developmental heterogeneity outcomes (Fig. 3b and Extended Data Fig. 2a). Unfortunately, females from these cohorts (Extended Data Fig. 2d) did not emerge as confidently separable bimodal groups (Extended Data Fig. 2e–g), nor did they show substantial cancer incidence (Extended Data Fig. 2h,i).

Trim28^{+/D9} males from the cohort thus satisfied the experimental design requirements for testing whether intrinsic developmental heterogeneity impacts cancer outcomes, namely, a reproducible developmental bifurcation and a spontaneous MCS. We therefore compared survival, tumor prevalence and associated tumor outcomes between light and heavy *Trim28*^{+/D9} male mice (Fig. 3d–j). Intriguingly, whereas 87% of *Trim28*^{+/D9}-heavy mice reached the endpoint free of illness, the majority of *Trim28*^{+/D9}-light mice showed aggressive MCS (Fig. 3d–j). Mean survival time was reduced in the *Trim28*^{+/D9}-light group, indicating a significantly accelerated disease course (57.4 and 68.5 weeks, respectively) (Fig. 3e). The only two *Trim28*^{+/D9}-heavy mice requiring euthanasia before the endpoint had bone marrow-derived acute myeloid leukemia (AML; Figs. 3e–j, top left and 2) and a set of four different primary carcinomas (Fig. 3e–i). *Trim28*^{+/D9}-light mice in the aggressive group exhibited up to four different primary tumors per mouse (Fig. 3h), mainly consisting of carcinomas, age-related carcinomas and benign tumors (Fig. 3f). Of these, malignant tumors were found in seminal vesicles, the prostate, the pancreas, lungs, the esophagus and the colon (Figs. 3g,i,j, top middle and right, bottom and 2). Consistent

with the differential timing of cancer onset between light and heavy mice, the endpoint analysis was dominated by tumors from *Trim28*^{+/D9}-heavy mice (Extended Data Fig. 3a–e). TRIM28 protein expression measures made at the endpoint in healthy, but cancer-prone, tissues (seminal vesicles, prostates, lungs, intestines and pancreata) showed no remarkable differences across genotypes and between *Trim28*^{+/D9}-light and *Trim28*^{+/D9}-heavy mice (Extended Data Figs. 3f and 4). Although these data do not exclude the possibility that TRIM28 is differentially expressed between morphs in a specific window of development, they argue against TRIM28 dosage effects as underpinnings for the observed differences in cancer outcomes. Thus, *Trim28*^{+/D9}-light developmental morphs exhibit an accelerated MCS. The data provide genetic evidence that TRIM28-dependent developmental heterogeneity controls later-life cancer outcomes.

Cancer risk morphs show unique early-life epigenetic states

We reasoned that if developmentally programmed epigenetic differences impact cancer susceptibility and outcomes later in life, then these differences should be detectable early in life. Therefore, we performed DNA methylation profiling on biopsies from the same male mice at 10 days of age (that is, before weaning). We used ear clips, as they are considered minimally invasive for young mice and similar to tissues used to detect early-life epigenetic signatures in humans^{29,30}. We used Illumina Infinium Mouse Methylation BeadChips to quantify DNA methylation at ~285,000 CpG sites including most annotated genes, functional RNAs and *cis*-regulatory regions of the mouse genome³¹. Global DNA methylation profiles were highly correlated across genotypes, indicating robust methodology (Fig. 4a, rows 1–4). Notably, samples containing the *Trim28*^{+/D9} allele (*Trim28*^{+/D9} and *Trp53*^{R270H/+}; *Trim28*^{+/D9}) exhibited approximately three times more differentially methylated CpG loci (DML) than those with *Trp53*^{R270H/+} alone when comparing to wild-type mice (Fig. 4b). DML in *Trim28*^{+/D9} and in *Trp53*^{R270H/+}; *Trim28*^{+/D9} double mutants overlapped strongly (Fig. 4a–d), indicating that *Trim28*^{+/D9} reproducibly changes the early-life methylome.

Relative to wild type, *Trim28*^{+/D9} biopsies showed an overall reduction in DNA methylation (Fig. 4e). This is consistent with the role of TRIM28 in gene silencing^{23,24}. Interestingly, *Trp53*^{R270H/+} mice also showed signs of early-life epigenetic changes, with some similarity to *Trim28*^{+/D9}-induced changes (Fig. 4f), consistent with evidence that p53 indirectly modulates DNA methylation levels^{32,33}. *Trim28*^{+/D9} hypomethylated DML were enriched in regions targeted by TRIM28, with probe set enrichment analysis revealing annotations for heterochromatin, monoallelic methylation, Polycomb silencing, CCCTC-binding

Fig. 3 | TRIM28-dependent developmental heterogeneity primes cancer outcomes.

a, Body mass distribution at 16 weeks by genotype. Each dot represents one male. Box plots show 25th, 50th (median) and 75th percentiles, with whiskers extending to 1.5 × interquartile range of the hinge and outliers plotted individually. Two-sided Levene’s test, Benjamini–Hochberg adjusted. Significance for adjusted *P* value (*P*_{adj}) < 0.05: WT versus *Trim28*^{+/D9}, *P*_{adj} = 0.0025152 (*F* = 13.482); WT versus *Trp53*^{R270H/+}, *P*_{adj} = 0.2031600 (*F* = 1.9399); WT versus *Trp53*^{R270H/+}; *Trim28*^{+/D9}, *P*_{adj} = 0.0067860 (*F* = 9.13); *Trim28*^{+/D9} versus *Trp53*^{R270H/+}, *P*_{adj} = 0.0025152 (*F* = 12.126); *Trim28*^{+/D9} versus *Trp53*^{R270H/+}; *Trim28*^{+/D9}, *P*_{adj} = 0.8220000 (*F* = 0.051); *Trp53*^{R270H/+}; *Trim28*^{+/D9} versus *Trp53*^{R270H/+}, *P*_{adj} = 0.0067860 (*F* = 8.5555). *n* = 130 male mice (16 WT, 35 *Trim28*^{+/D9}, 41 *Trp53*^{R270H/+}, 38 *Trp53*^{R270H/+}; *Trim28*^{+/D9}). **b**, Fat and lean mass data at 16 weeks by genotype, with overlaid mclust density estimation. Each dot represents one male; dot size is proportional to mclust classification certainty. *n* = 138 male mice (18 WT, 36 *Trim28*^{+/D9}, 43 *Trp53*^{R270H/+}, 41 *Trp53*^{R270H/+}; *Trim28*^{+/D9}). **c**, Bimodality index: ratio of mclust-determined Bayesian information criterion (BIC) for two clusters over one from fat and lean mass data at 16 weeks. *n* = 138 male mice (same as in **b**: 18 WT, 36 *Trim28*^{+/D9}, 43 *Trp53*^{R270H/+}, 41 *Trp53*^{R270H/+}; *Trim28*^{+/D9}). **d**, Proportion of *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light males with malignant aggressive tumors as a percentage of males per group (before and after endpoint). Actual numbers are given in fractions. Two-sided two-sample test for equality of proportions without continuity correction; significance for *P* < 0.05: heavy versus light, *P* = 0.01466,

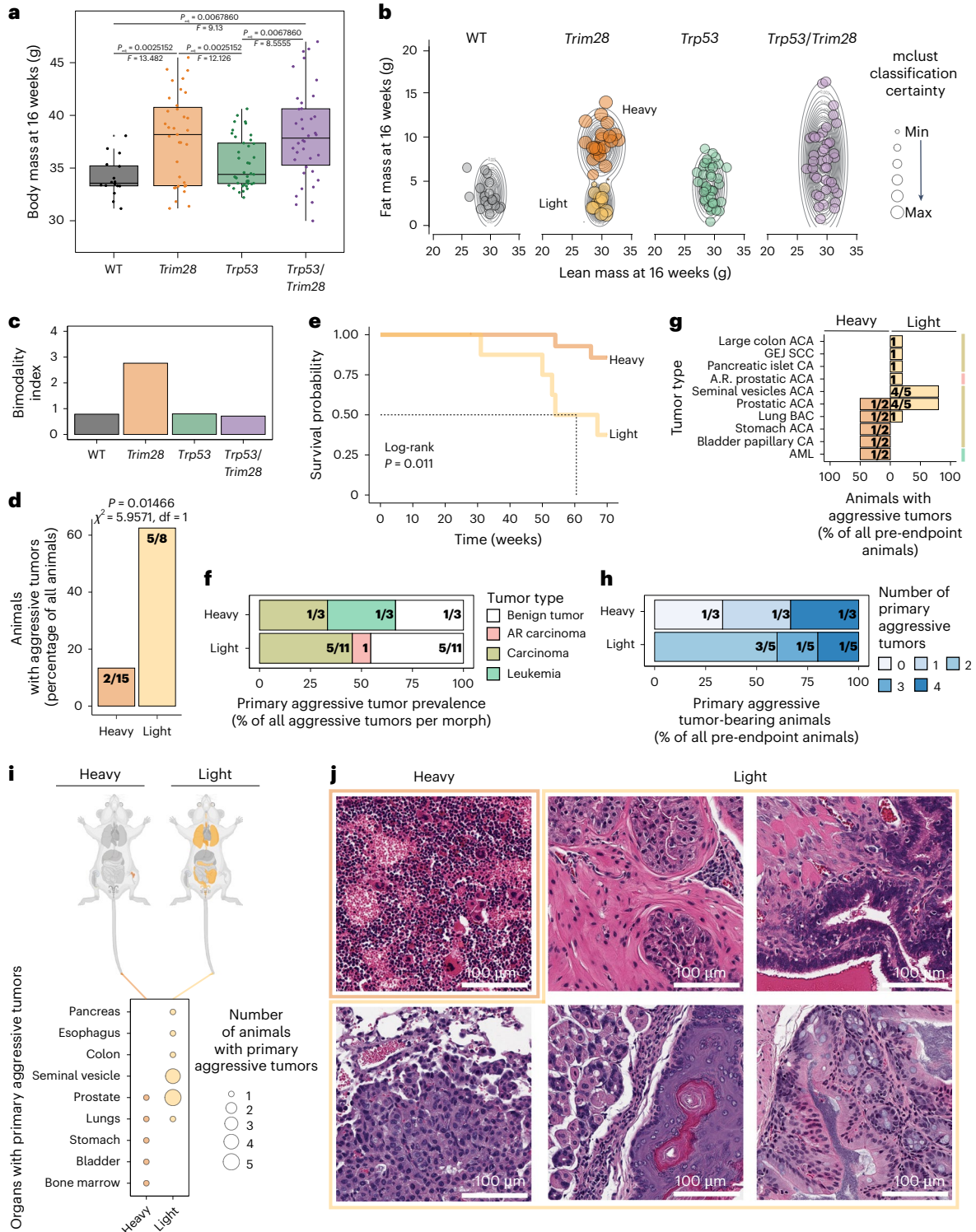
$\chi^2 = 5.9571$, degrees of freedom (df) = 1. *n* = 23 male mice (15 heavy, 8 light). **e**, Kaplan–Meier survival probability of *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light mice. Log-rank test, *P* = 0.011. *n* = 23 males (15 heavy, 8 light). **f**, Prevalence of aggressive tumor types in *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light mice as a percentage of total aggressive tumors per group. Actual numbers are shown in fractions. *n* = 14 tumors (malignant and benign; 3 in heavy and 11 in light). **g**, Distribution of tumor types in *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light mice, as a percentage of mice with malignant aggressive tumors per group. Actual numbers are shown in fractions. Right, age-related carcinoma (red), carcinoma (gold), leukemia (light green). CA, carcinoma; GEJ, gastro-esophageal junction. *n* = 7 males (2 heavy, 5 light). **h**, Fraction of *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light mice with 0 or multiple malignant aggressive tumors. Actual numbers are shown in fractions. *n* = 8 males (3 heavy, 5 light). **i**, Tissues targeted by malignant aggressive tumors by genotype. Top, mouse anatomy plots; nontargeted tissues are in light gray, and targeted tissues are colored by genotype: WT (black), *Trim28*^{+/D9} (orange), *Trp53*^{R270H/+} (green), *Trp53*^{R270H/+}; *Trim28*^{+/D9} (purple). Bottom, organs with malignant aggressive tumors by morph. *n* = 7 males (2 heavy, 5 light). **j**, Histological examples of aggressive malignant tumors in *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light mice. Top left, AML, heavy (as in Fig. 2). Top middle, prostatic ACA, light. Top right, seminal vesicle ACA, light. Bottom left, BAC, light. Bottom middle, gastro-esophageal junction SCC, light. Bottom right, colon ACA, light. *n* = 7 males (2 heavy, 5 light). Panel **i**, top created with BioRender.com.

factor (CTCF), TRIM28 binding and histone H3 Lys9 trimethylation (H3K9me3) (Fig. 4g). Thus, TRIM28 dosage is required to maintain the fidelity of early-life DNA methylation patterns at heterochromatic regions.

Importantly, DNA methylation differences were also detected between *Trim28*^{+/D9} mice that would later go on to become light versus heavy, despite the isogenicity of these groups and the lack of detectable phenotypic differences at this early age (Figs. 5a–c and 4a, rows 5 and 6). Specifically, we found 1,133 DML between the two *Trim28*^{+/D9} morphs, including a clear skew toward relative hypomethylation in

the light, cancer-prone morph (Fig. 5a,b). This signature was partially maintained in three of four tissues when tested later in life (Fig. 5d). Based on correlation, the light morphs were technically more similar to wild-type mice (Fig. 4a). Furthermore, we found a specific correlation between the average level of methylation of light-specific hypomethylated probes and 16-week fat mass content in *Trim28*^{+/D9} mice but not in other genotypes (Fig. 5e,f). This indicates that the light-specific signature is not an obesity-related signature.

Early-life *Trim28*^{+/D9}-light hypomethylated DML were detected preferentially at transcribed (transcribed and weakly transcribed



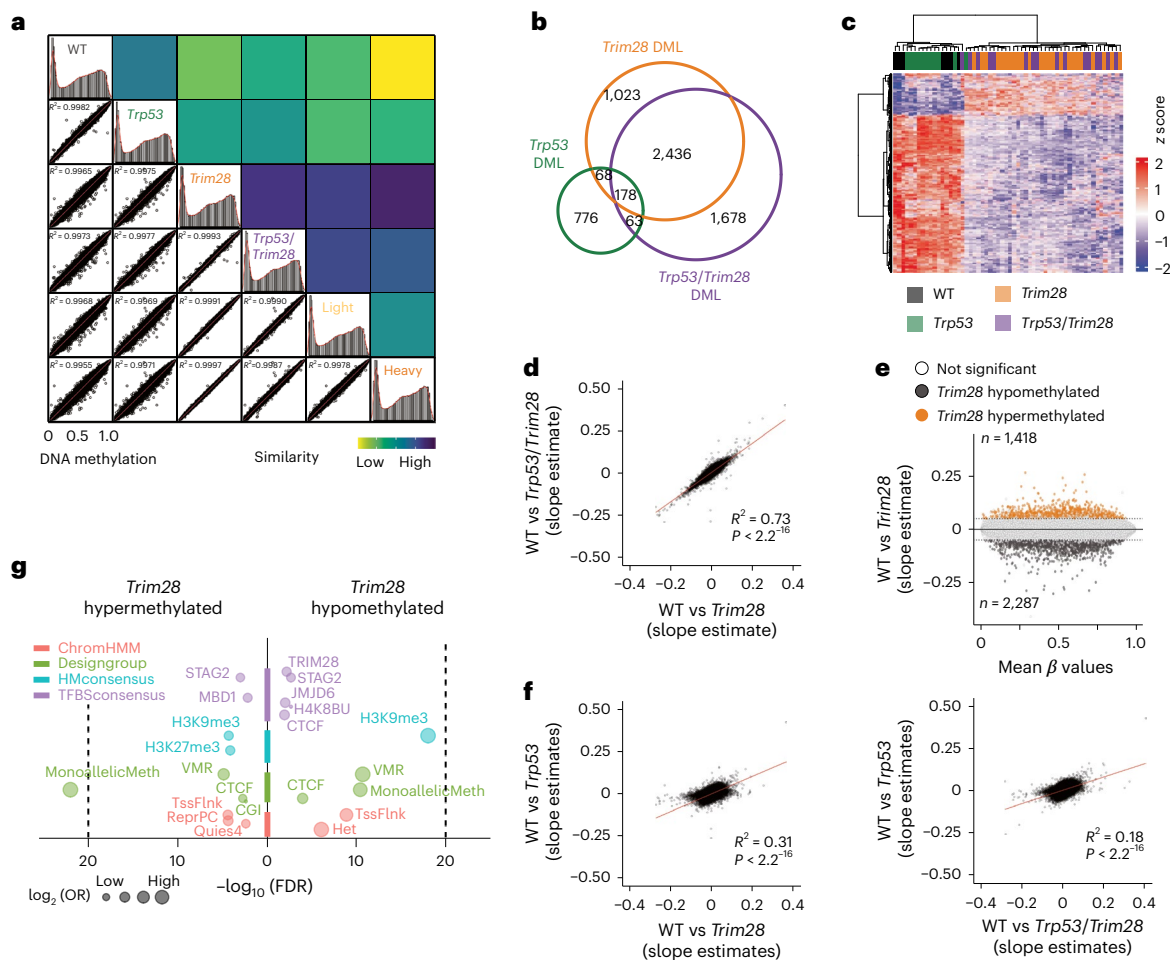


Fig. 4 | *Trim28*^{+/D9} effects on early-life epigenomes. a, Right, correlation plots comparing global DNA methylation on all available probes among all genotypes and *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light morphs as the sum of squared residuals from linear regressions. Left, density plots of β -value distribution for all genotypes and trajectories. R^2 values derive from Pearson correlations. $n = 58$ (7 WT, 24 *Trim28*^{+/D9}, 11 *Trp53*^{R270H/+}, 16 *Trp53*^{R270H/+}; *Trim28*^{+/D9}). **b**, Overlap of DML between *Trim28*^{+/D9}, *Trp53*^{R270H/+}; *Trim28*^{+/D9}, *Trp53*^{R270H/+} and WT mice. $n = 58$ (7 WT, 24 *Trim28*^{+/D9}, 11 *Trp53*^{R270H/+}, 16 *Trp53*^{R270H/+}; *Trim28*^{+/D9}). **c**, Heatmap of the z score of log_{1p}-transformed β values of differentially methylated probes in all genotypes relative to WT mice. Effect size cutoff = 0.1; P -value cutoff = 0.05 by t -testing slope estimates. $n = 58$ (7 WT, 24 *Trim28*^{+/D9}, 11 *Trp53*^{R270H/+}, 16 *Trp53*^{R270H/+}; *Trim28*^{+/D9}). **d**, Correlation plot comparing DNA methylation levels (slope estimate) on all available probes in WT versus *Trp53*^{R270H/+}; *Trim28*^{+/D9} and WT versus *Trim28*^{+/D9} mice. R^2 value = 0.73; $P < 2.2 \times 10^{-16}$ (two-sided Pearson correlation). $n = 58$ (7 WT, 24 *Trim28*^{+/D9}, 11 *Trp53*^{R270H/+}, 16 *Trp53*^{R270H/+}; *Trim28*^{+/D9}). **e**, M (slope intercept) A (average mean β -value) plot of differential DNA methylation levels (slope

estimate) in WT versus *Trim28*^{+/D9} mice. $n = 31$ (7 WT, 24 *Trim28*^{+/D9}). Black and orange dots represent hypomethylated and hypermethylated probes in *Trim28*^{+/D9} mice, 2,287 and 1,418 probes respectively. White dots are nonsignificant DML (estimate and P -value cutoff = 0.05). **f**, Correlation plots comparing DNA methylation levels (slope estimate) on all available probes in WT versus *Trp53*^{R270H/+} (left) and WT versus *Trp53*^{R270H/+}; *Trim28*^{+/D9} mice (right). R^2 values = 0.31 and 0.18; $P < 2.2 \times 10^{-16}$ (two-sided Pearson correlation). $n = 58$ male mice (7 WT, 24 *Trim28*^{+/D9}, 11 *Trp53*^{R270H/+}, 16 *Trp53*^{R270H/+}; *Trim28*^{+/D9}). **g**, Enrichment plots of features for *Trim28*^{+/D9}-dependent differentially methylated probes. H3K27me3, histone H3 Lys27 trimethylation; H4K88BU, histone H4 Lys8 butyrylation; Het, heterochromatin; MonoallelicMeth, monoallelic methylation; OR, odds ratio; Quies, quiescent gene; ReprPC, repressed by Polycomb; TssFlnk, flanking active TSS; VMR, variably methylated regions. False discovery rate (FDR) cutoff = 0.01 (one-tailed Fisher's exact test). $n = 58$ male mice (7 WT, 24 *Trim28*^{+/D9}, 11 *Trp53*^{R270H/+}, 16 *Trp53*^{R270H/+}; *Trim28*^{+/D9}).

states, 6% of light- and 1% of heavy-specific probes; Fig. 6a) and active and bivalent (72% versus 55% of probes, respectively; Fig. 6b) transcriptional start sites (TSS) as well as at CpG islands (CGIs; 14% versus 10% of probes) and shelf regions (4% versus 1% of probes). *Trim28*^{+/D9}-heavy mice, by contrast, were hypomethylated at repressed weak Polycomb domains (14% versus 2% of probes; Fig. 6a) and regions flanking the TSS and TSS-distal regions (>150 kb from the TSS, 14.4% versus 8.7%; Fig. 6c). Of note, differential CGI methylation preferentially impacted intermediately methylated CGIs and not canonical unmethylated CGIs (Fig. 6d). Essentially no differential DNA methylation was detected in canyons^{34,35} (Fig. 6e), suggesting that these regions are refractory to the TRIM28 early-life heterogeneity effects. Heavy-specific hypomethylation was also enriched at genic and strong enhancers (47% versus 26% of probes), while light-specific

hypomethylation was enriched at poised and weak enhancers (57% versus 40% of probes) (Fig. 6f). Thus, *Trim28*^{+/D9} cancer susceptibility morphs exhibit early-life epigenome differences at repressive and cis-regulatory regions.

Given the role of TRIM28 in regulating repetitive elements³⁶, we also checked these regions. Interestingly, light mice showed specific hypomethylation at DNA repeats, satellites and transfer RNAs. *Trim28*^{+/D9}-heavy mice, instead, showed relative hypomethylation at retrotransposons (29% and 41% of probes on long interspersed nuclear elements and long terminal repeats in heavy mice, compared with 18% and 25% in light mice) (Fig. 6g).

Probe set enrichment analysis revealed differential methylation predominantly at regions of monoallelic methylation and imprinting, including the *Kcnq1-Kcnq1ot1* cluster, *H19* and *Peg3* (Fig. 7a, top, green

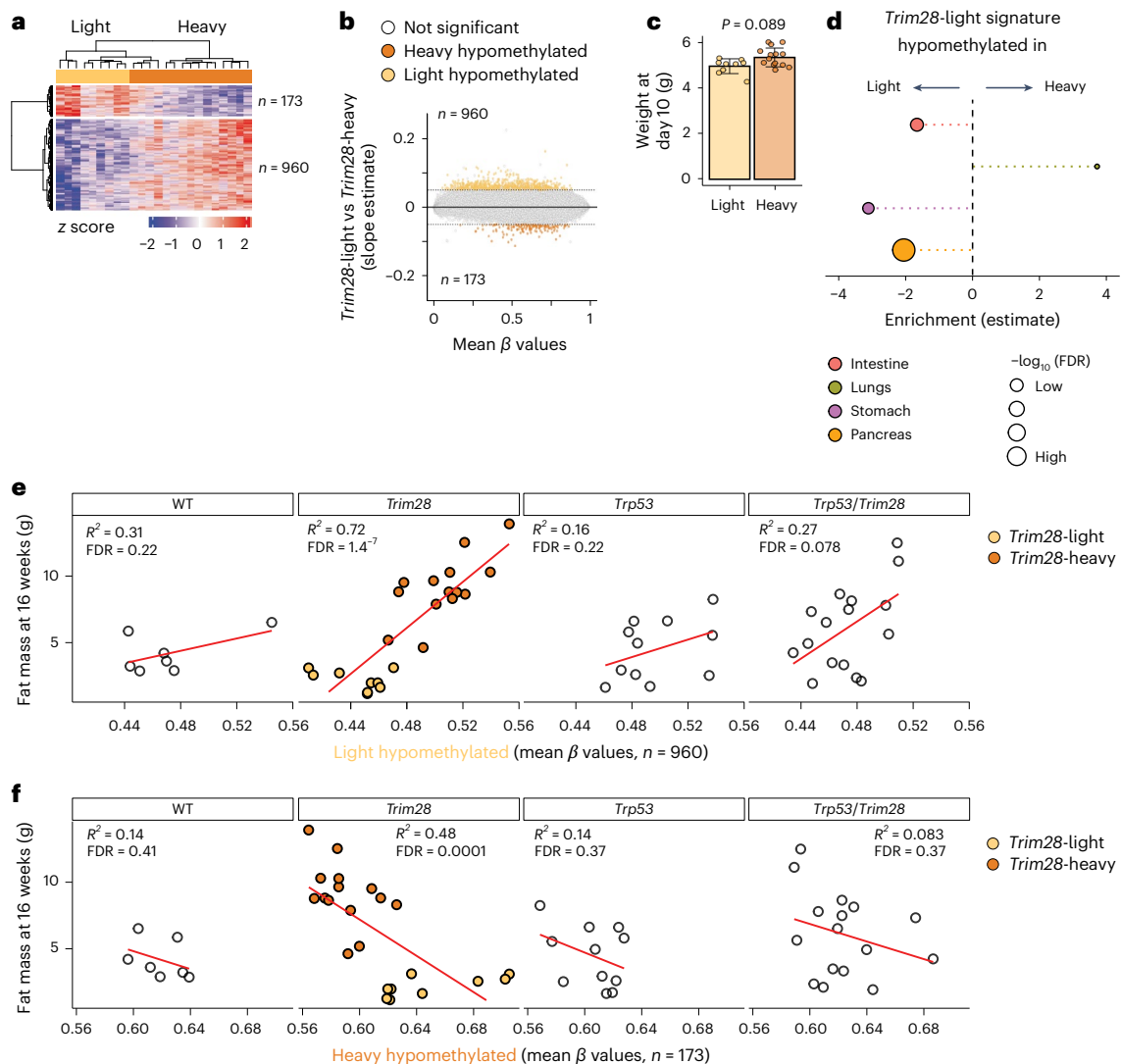


Fig. 5 *Trim28*^{+/D9}-dependent cancer susceptibility states are distinguished by distinct early-life epigenomes. **a**, Heatmap reporting z-score transformed β values of differentially methylated probes in *Trim28*^{+/D9}-heavy versus *Trim28*^{+/D9}-light mice. Effect size cutoff = 0.05; P-value cutoff = 0.001 by *t*-testing the slope estimates. n = 24 male mice (15 *Trim28*^{+/D9}-heavy, 9 *Trim28*^{+/D9}-light). **b**, MA plot of differential DNA methylation levels (slope estimate) in *Trim28*^{+/D9}-heavy versus *Trim28*^{+/D9}-light male mice. Dark and light orange dots represent hypermethylated and hypomethylated probes in *Trim28*^{+/D9}-light mice, respectively (estimate and P-value cutoff = 0.05 by *t*-testing the slope estimates). White dots are not significant. n = 24 male mice (15 heavy, 9 light). **c**, Weight at day 10 of *Trim28*^{+/D9}-light and *Trim28*^{+/D9}-heavy male mice used for DNA methylation

array experiments. n = 24 male mice (15 heavy, 9 light). P value = 0.089 (unpaired two-sample Wilcoxon test). **d**, Enrichment plot of *Trim28*^{+/D9}-light hypomethylated probes in *Trim28*^{+/D9}-light versus *Trim28*^{+/D9}-heavy male mice in the indicated tissues. FDR cutoff = 0.01 (one-tailed Fisher's exact test). n = 46 tissues (12 pancreata, 12 lungs, 11 stomachs, 11 intestines). **e, f**, Scatterplots showing the correlation by genotype between the average level of DNA methylation on the indicated probe sets (*Trim28*^{+/D9}-light or *Trim28*^{+/D9}-heavy hypomethylated probes) and fat mass at 16 weeks. Colors indicate the morphs: *Trim28*^{+/D9}-light (light orange) and *Trim28*^{+/D9}-heavy mice (dark orange). n = 58 male mice (7 WT, 24 *Trim28*^{+/D9}, 11 *Trp53*^{R270H/+}, 16 *Trp53*^{R270H/+}; *Trim28*^{+/D9}).

and bottom, light orange) and regions annotated as decorated with H3K9me3 and H3K27me3 (Fig. 7a, top, light blue). A search for overlap with transcription factor binding (Fig. 7a, top, purple) revealed strong and specific enrichment for DNA methylation-binding proteins (MBD1, MECP2, C17orf96, DPPA2 and TRIM28 itself) and Polycomb-silencing machinery (SUZ12, EZH2, C17orf96, RNF2, AEBP2, PCGF2, CBX7, BMI1 and JARID2). Enrichment was also observed for probes within the epigenetic aging clock (Fig. 7a, top, green). Representative DML between *Trim28*^{+/D9}-light and *Trim28*^{+/D9}-heavy mice are shown in Extended Data Fig. 5a–d. Collectively, these data suggest that *Trim28*^{+/D9}-light mice have more permissive chromatin at regions that would otherwise be silenced. Thus, early-life methylation signatures can distinguish developmentally programmed cancer susceptibility states.

TRIM28 silences oncogenes early in life

Interestingly, genes specifically hypomethylated in *Trim28*^{+/D9}-light mice were also enriched for epigenetic regulators of gene expression, heterochromatin formation, heterochromatin organization, genomic imprinting, DNA methylation and DNA alkylation (Fig. 7a, bottom, light orange and Fig. 7b). Thus, light, cancer-prone *Trim28*^{+/D9} morphs exhibit hypomethylation at coding regions of epigenetic silencers and at their targeted genomic regions (Fig. 7a, top left).

This light-specific hypomethylated gene signature was globally downregulated (Extended Data Fig. 5e), which correlates with its hypermethylated state in *Trim28*^{+/D9} mice compared with wild-type mice (Extended Data Fig. 5f). RNA sequencing (RNA-seq) of matched healthy and tumor tissues from cancer-susceptible organs in the *Trim28*^{+/D9} model revealed that *Trim28*^{+/D9}-light hypomethylated genes were

globally upregulated in tumors, gene- and tissue-specific effects notwithstanding (Extended Data Fig. 5g,h). These genes were also differentially expressed in *Trim28^{+/D9}*-light versus *Trim28^{+/D9}*-heavy tumors (Extended Data Fig. 5i), suggesting involvement in defining the two developmental cancer susceptibility states. We did not observe any difference in mutational rate in *Trim28^{+/D9}*-light hypomethylated genes, comparing light and heavy mice (Extended Data Fig. 5j).

Given that later-life metabolic phenotypes distinguish the two developmental morphs, we tested for dysregulation of metabolic pathways. Of 674 annotated metabolic pathways, only 5 showed evidence of pathway-level change. Specifically, *Trim28^{+/D9}*-light hypomethylated DML showed modest enrichment in nucleobase-containing metabolic pathways (Extended Data Fig. 6a). The data are consistent with the lack of detectable early-life metabolic differences between morphs. Thus, epigenome rewiring occurs before observable metabolic differences between *Trim28^{+/D9}* morphs.

Next, we asked whether the observed early-life dysregulation was likely to impact cancer outcomes. Querying the Jensen DISEASES database³⁷, we found enrichment of imprinting disorders and cancer associations in light versus heavy differentially methylated ‘signature’ genes (Extended Data Fig. 6b). Similarly, querying the COSMIC Cancer Gene Census, a database of high-confidence human oncogenes and tumor-suppressor genes³⁸, we found that *Trim28^{+/D9}*-light hypomethylated genes were significantly enriched for known human oncogenes (*GNAS*, *JAK3*, *MYCN* and *HMG2*) (Fig. 7c). Also, focusing on The Cancer Genome Atlas (TCGA) Pan-Cancer Atlas³⁹, we found that patients with mutations in *TRIM28^{+/D9}*-light hypomethylated signature genes showed reduced overall survival probability (Fig. 8a, left) and a striking difference in time to relapse (Fig. 8a, right) compared to patients bearing other mutations. This was true for male and female patients (Extended Data Fig. 6c,d). The same analyses of orthologs for hypermethylated genes in light (hypomethylated in heavy) mice showed no significant differences (Extended Data Fig. 6e). Stratifying the data by tumor type showed widespread tumor-promoting effects in orthologs of essentially all *TRIM28^{+/D9}*-light signature genes, with significantly reduced survival rates across many tumor types (Fig. 8b and Extended Data Fig. 6f) and sex-specific effects of distinct mutations (Extended Data Fig. 6g,h). Of note, prostatic adenocarcinoma (ACA), the most prevalent cancer type in *Trim28^{+/D9}*-light mice (Fig. 3g,i and Extended Data Fig. 3b,d), showed cancer-accelerating associations with nearly all human orthologs of light-specific signature genes (Fig. 8b, leftmost column and Extended Data Fig. 6f,h). A tissue-specific effect was also observed when looking at the correlation between gene expression or DNA methylation at TSSs of *Trim28^{+/D9}*-light signature genes and survival (Extended Data Fig. 7). Well-known oncogenes showed the strongest and most consistent effects. For example, *HMG2* expression was consistently associated with poor overall and disease-free survival across multiple cancer types. Consistent with canonical DNA methylation-dependent silencing at TSSs, the DNA methylation level at the *HMG2* TSS was positively

associated with survival. This reciprocal pattern of correlation was unequivocal across multiple cancers including head and neck squamous cell carcinoma (SCC), kidney renal clear cell carcinoma, pancreatic ACA and mesothelioma (Extended Data Fig. 7). Of note, germ cell tumors contained the most correlations between DNA methylation levels and overall survival (Extended Data Fig. 7). This type of tumor is known to be associated with globally demethylated DNA⁴⁰. Finally, we also tested for mutational co-occurrence, a phenomenon that can highlight potential additive or synergistic potential among mutations. Consistent with the phenotype of the *Trim28^{+/D9}*-light morphs, co-occurrence of light signature gene mutations was markedly overrepresented in human primary tumors (Fig. 8c). Thus, the early-life *Trim28^{+/D9}*-dependent epigenome regulates bona fide oncogenes.

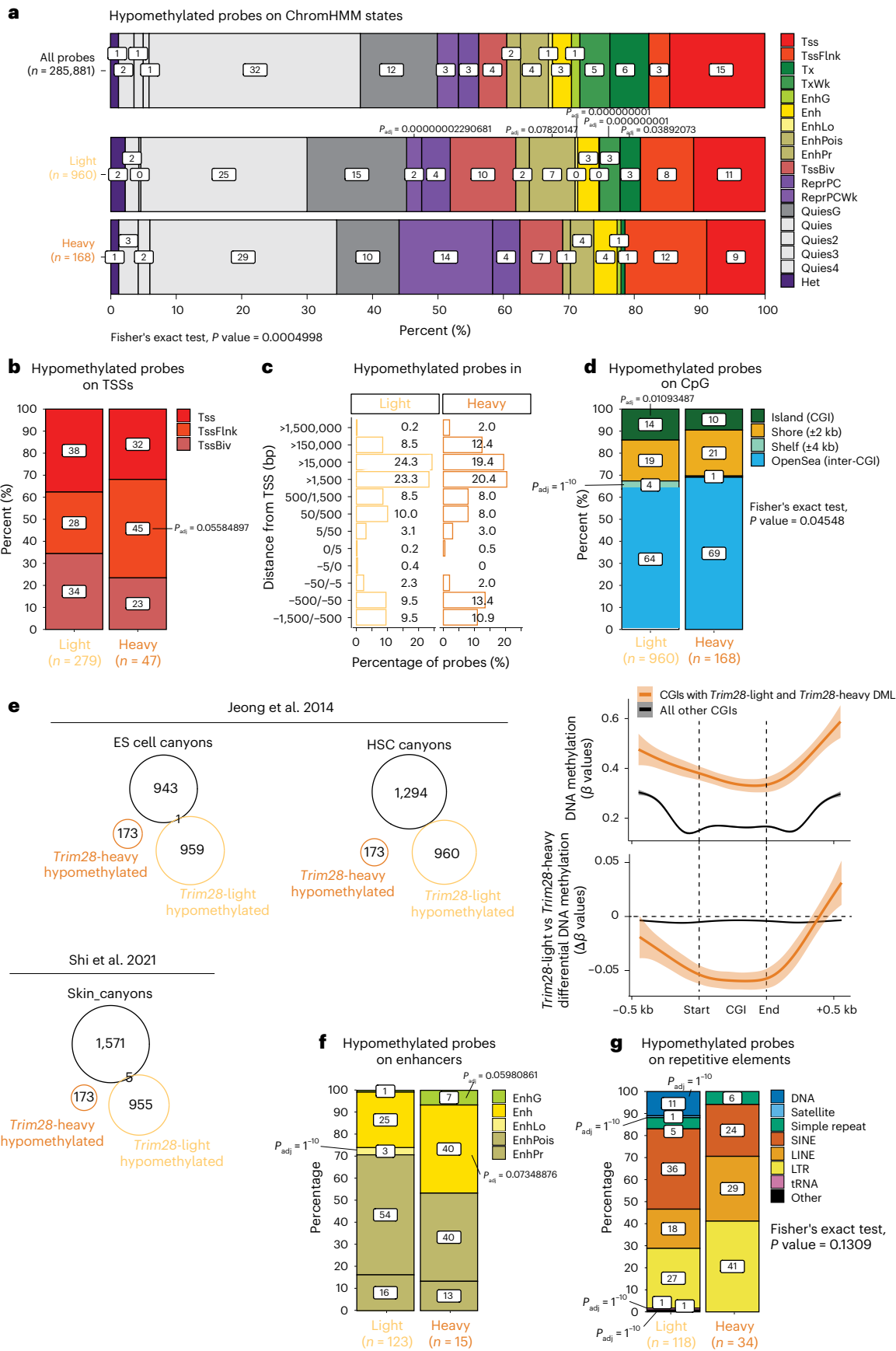
Overall, these data identify *Trim28^{+/D9}*-light signature genes as putative mediators of the altered cancer susceptibility states found in *Trim28^{+/D9}* male mice. They suggest a model in which intrinsic differences in early-life epigenetic programming influence cancer outcomes (Fig. 8d).

Discussion

Here, we demonstrate that (TRIM28-buffered) intrinsic developmental heterogeneity induces an epigenetically defined developmental bifurcation early in life, which is linked to distinct cancer susceptibility later in life. Specifically, TRIM28 haploinsufficiency generates two reproducible developmental morphs (at the organismal level) that differ in their cancer susceptibility: one relatively ‘prone’ and one ‘resistant’, where these two categories correspond to accelerated and delayed tumor development. This idea has parallels to epigenetic heterogeneity effects described within tumors and between tumors, except at the interorganismal level. How meta-stable states between identical cells or organisms are imposed remains unclear, although pioneering work with variegating reporters implicates epigenetic silencing machinery^{41,42}. It has been suggested that a key condition for the emergence of alternate cellular states is the epigenetic reorganization of the genome^{43,44}. Feinberg and Levchenko²⁰ provided a theoretical framework for how genetic and epigenetic networks might generate meta-stable functional states across cells, a potential energy landscape model that includes energy wells or ‘attractors’. In that model, DNA mutations and/or changes in epigenetic topology (for example, DNA methylation or histone modifications) can alter that landscape and create alternate attractor states. Our data provide evidence that these concepts hold true at the organismal scale and that these differences can have real consequences for carcinogenesis. They suggest that TRIM28-dependent silencing helps define the shape of the potential energy landscape (for example, by controlling the depth of or barrier between attractor states). In the same way that oncogenic mutations have different effects depending on the cellular developmental stage^{8,45}, our data suggest that oncogenesis can also be influenced by organism-wide epigenetic programs that are established in development.

Fig. 6 | Characterization of *Trim28^{+/D9}*-light and *Trim28^{+/D9}*-heavy differentially methylated probes. a, Stacked bar plots showing the distribution of hypomethylated probes from *Trim28^{+/D9}*-light and *Trim28^{+/D9}*-heavy male mice across ChromHMM states. Statistical significance was assessed using a two-sided Fisher’s exact test with a simulated *P* value (based on 2,000 replicates); *P* = 0.0005. Asterisks mark regions enriched in either *Trim28^{+/D9}*-light or *Trim28^{+/D9}*-heavy hypomethylated probes, with statistical significance determined by a hypergeometric test and FDR correction (*P*_{adj} < 0.1). The overall distribution of all array probes is included for comparison. *n* = 24 male mice (15 heavy, 9 light). Tss, active TSS; Tx, transcription; TxWk, weak transcription; EnhG, genic enhancer; Enh, enhancer; EnhLo, weak enhancer; EnhPois, poised enhancer; EnhPr, primed enhancer; TssBiv, bivalent TSS; ReprPCWk, repressed by Polycomb, weak. **b**, Stacked bar plots display the distribution of hypomethylated probes across TSSs, with adjusted *P* values indicating regions enriched in *Trim28^{+/D9}*-light and *Trim28^{+/D9}*-heavy probes. Statistical significance was determined by a one-sided

hypergeometric test with FDR correction (*P*_{adj} < 0.1). *n* = 24 male mice (15 heavy, 9 light). **c**, Bar plots showing the distribution of *Trim28^{+/D9}*-light and *Trim28^{+/D9}*-heavy hypomethylated probes relative to TSSs. *n* = 24 male mice (15 heavy, 9 light). **d, f, g**, Stacked bar plots representing the distribution of hypomethylated probes across enhancers (**d**), CGIs (**f**) and repetitive elements (**g**). SINE, short interspersed nuclear element; LINE, long interspersed nuclear element; LTR, long terminal repeat; tRNA, transfer RNA. Statistical significance was assessed using Fisher’s exact test with a simulated *P* value and the hypergeometric test with FDR correction (*P*_{adj} < 0.1). *n* = 24 male mice (15 heavy, 9 light). **d**, Bottom, DNA methylation levels in *Trim28^{+/D9}*-light and *Trim28^{+/D9}*-heavy male mice at CGIs with DML (orange) and other CGIs (black), presented with linear model smooth lines and 0.95 confidence intervals (‘lm’ method). **e**, Overlap of hypomethylated probes from *Trim28^{+/D9}*-light and *Trim28^{+/D9}*-heavy mice with annotated DNA methylation canyons in embryonic stem (ES) cell, hematopoietic stem cells (HSCs)³⁴ and mouse skin³⁵. *n* = 24 male mice (15 heavy, 9 light).



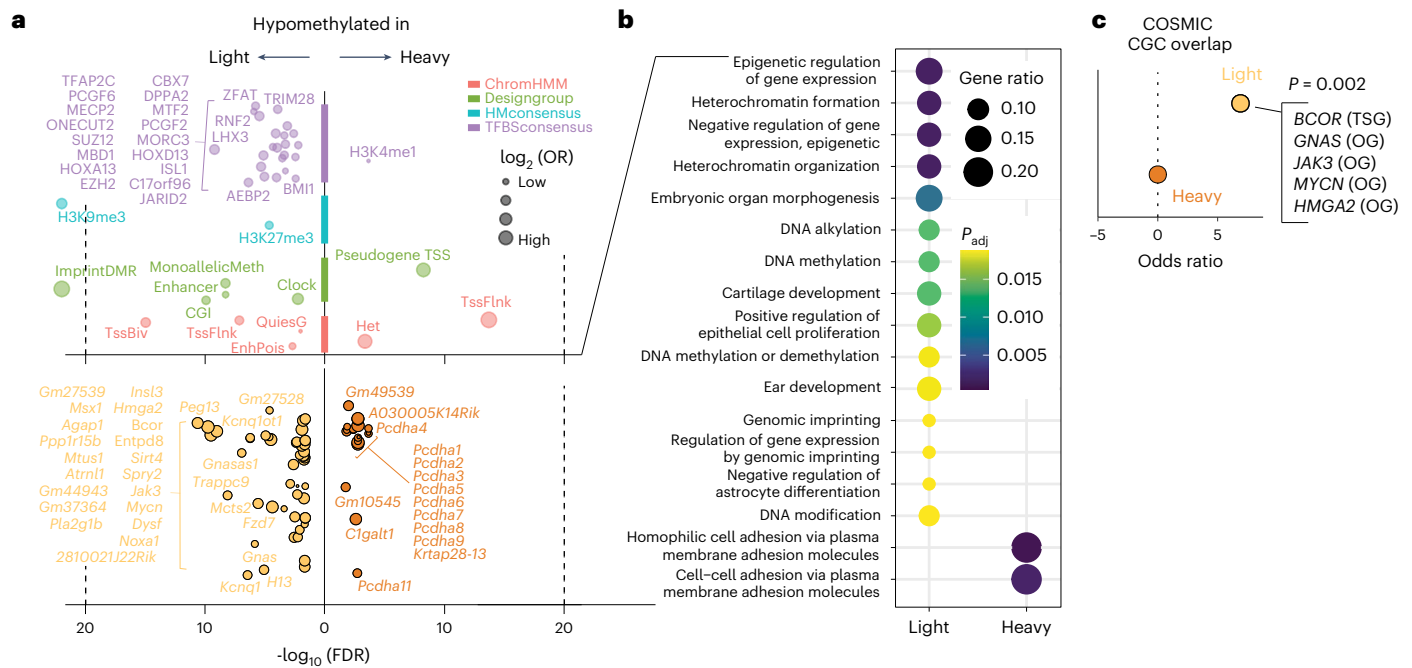


Fig. 7 | *Trim28*^{+/D9}-dependent early-life epigenomes are enriched for epigenetic regulators and bona fide oncogenes. a, Enriched features (top) and genes (bottom) for probes differentially methylated between *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light mice. H3K4me1, histone H3 Lys4 monomethylation. Bottom, the number of genes found differentially methylated is reported for *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light mice. FDR cutoff = 0.01 (one-tailed Fisher's exact test). **b**, Gene Ontology enrichment plot of biological processes enriched in DML between *Trim28*^{+/D9}-heavy versus *Trim28*^{+/D9}-light mice. The size of the dots represents the gene ratio, while the color represents the P_{adj} value for each term;

Fisher's exact test P value with Benjamini–Hochberg adjustment for multiple comparisons, cutoff = 0.05. The number of total genes analyzed in each group is reported. **c**, Odds ratio from one-sided Fisher's exact tests, testing the statistical significance of enrichment between genes enriched by hypomethylated probes in *Trim28*^{+/D9}-heavy (dark orange) and *Trim28*^{+/D9}-light mice (light orange) and the COSMIC Cancer Gene Census (CGC; <https://cancer.sanger.ac.uk/census>). OG, oncogene; TSG, tumor-suppressor gene. $P = 0.0002$. In **a–c**, $n = 24$ male mice (15 *Trim28*^{+/D9}-heavy, 9 *Trim28*^{+/D9}-light).

This study also bridges two key questions in cancer epigenetics: how does cellular epigenetic state dictate oncogenic transformation^{8–10,46–48}, and how do early-life epigenetic cues modulate cancer risk^{21,22,46}? Our data suggest that interindividual differences in early-life epigenome organization influence differential cancer development, prevalence and survival. The interindividual epigenetic differences identified in this study, alongside prior work indicating that H3K9me3 strongly correlates with tumor mutation density⁴⁹, suggest that one potential mechanism for the observed differential cancer outcomes between morphs is an altered resistance to mutation. Once tumors were established, we could not detect significant differences in mutations between the two morphs, arguing against a model where TRIM28 buffers against a small number of very specific hotspot mutations. We also detected important differences in Polycomb-targeted genes (which have been recently associated with cancer-prone cell fates¹¹) and poised and bivalent regions between the cancer susceptibility morphs. During tumorigenesis, these regions are sensitive to regulation by DNA methylation and correlate with cell 'stemness' (refs. 47,48,50). Regardless, the provocative implication from our data is that individual cancer susceptibility may have as much to do with the epigenetic 'background' we are born with as it does with DNA mutation, external environment and cell of origin. Just as prior work demonstrated that epigenetic dysregulation at specific genes drives tumorigenesis in specific tissues and cell differentiation stages, we would expect that the DML identified here trigger tissue-specific and developmental stage-specific effects. Key questions to understand include when the distinct cancer-susceptible epigenetic backgrounds become 'activated' in the *Trim28*^{+/D9} line, why some tissues are more sensitive than others and how sensitive these effects are to genetic background. A previous report indicated parent-of-origin-specific

effects of TRIM28 in regulating imprinting⁵¹. A priority comparison for the future will therefore be to evaluate whether inheritance of *Trim28* haploinsufficiency from one or the other parent triggers differing outcomes.

Of note, our findings are distinct from the published TRIM28 oncogene and tumor-suppressor functions²⁵ that derive from experiments using complete (homozygous) knockouts. In contrast to homozygous knockout models, the *Trim28*^{+/D9} mouse exhibits near normal levels of TRIM28 (ref. 27). Given the presence of TRIM28 in several complexes (TRIM28 co-repressor complex, NuRD, CoREST, PML-NB, BORG–TRIM28, ZMYM2–TRIM28, MAGE–TRIM28, HUSH or ATM–CHK2 and ATR–CHK1 (refs. 52)), this is an important distinction between the models. The DNA methylation differences between light and heavy morphs suggest that *Trim28*^{+/D9} specifically affects the TRIM28 silencing function. This fundamental difference between models may explain for instance why *Trim28*-knockout models develop liver tumors⁵³ and *Trim28*^{+/D9} mice do not.

Our work provides genetic evidence that intrinsic developmental heterogeneity influences overall cancer risk. The ability to show this effect in multiple tissue types is both a strength and a limitation of this study. By focusing on an MCS model, we demonstrate that differential susceptibility is a property of the entire organism and can identify responsive tissues. Thousands of animals would, however, be needed to draw the same conclusions for all the observed cancer subtypes including rare cancers. Likewise, and because the *Trim28*^{+/D9} mutation in this model occurs in the whole body, it will be difficult to use the *Trim28*^{+/D9} mouse (by itself) to dissect the molecular mechanisms that underlie each observed cancer. A natural extension therefore is to refine the model to understand the mechanistic basis of developmentally primed cancer susceptibility for each cancer type.

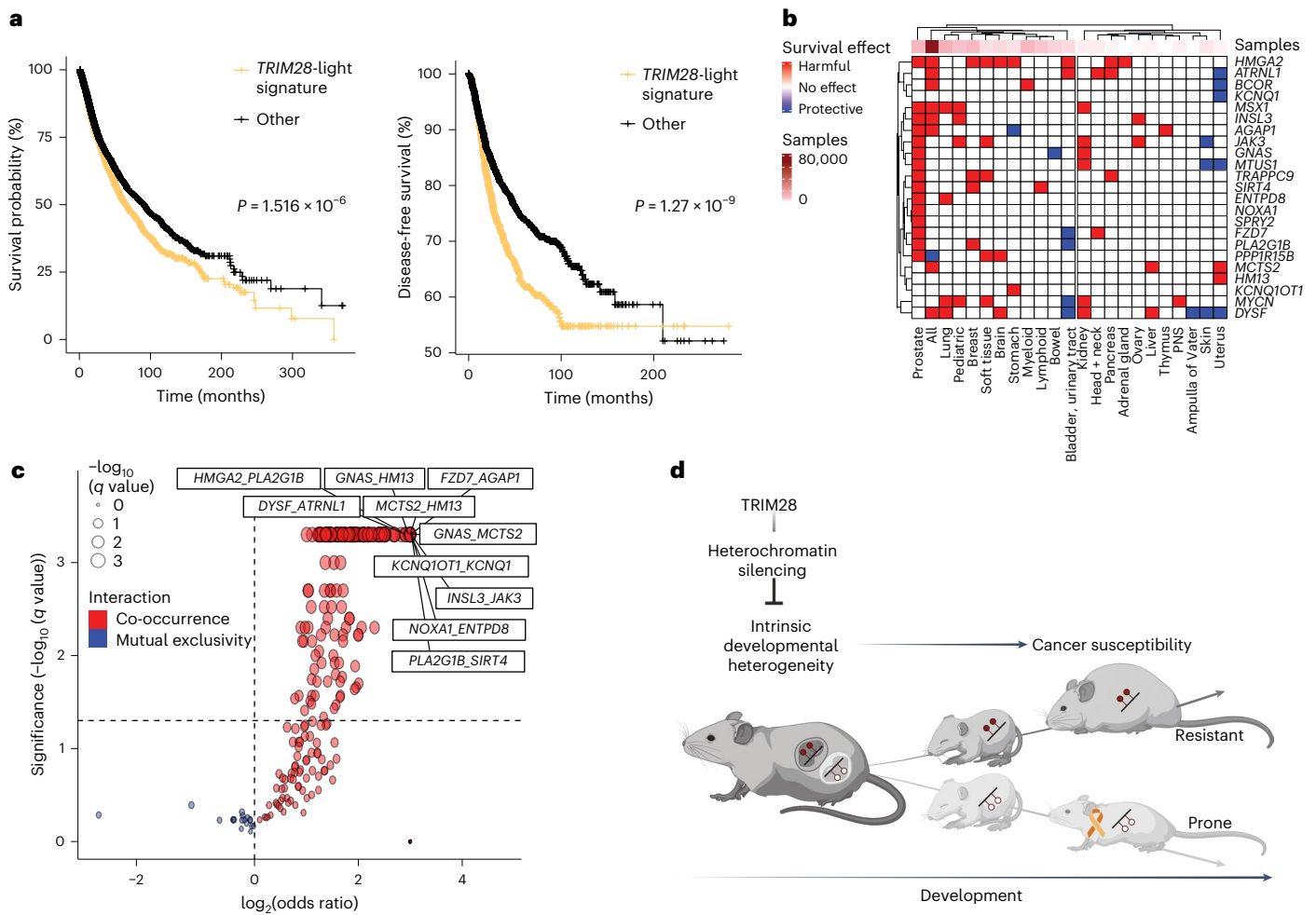


Fig. 8 *Trim28^{R270H}*-light hypomethylated signature genes are associated with reduced survival probability in humans. **a**, Left, Kaplan–Meier survival probability as a percentage of the total population and time of survival in months. Log-rank test, $P = 1.516 \times 10^{-6}$. Total samples analyzed, 10,967. Right, Kaplan–Meier disease-free survival probability. Log-rank test, $P = 1.27 \times 10^{-9}$. Left and right, all TCGA Pan-Cancer Atlas patients with mutations in genes from the *TRIM28^{R270H}*-light hypomethylated signature (light orange, $n = 3,766$ samples) are compared to individuals without mutations in the same genes (black, $n = 7,180$ samples). **b**, Heatmap of the effects on overall survival probability of mutations in the indicated genes and tumor tissues. The analysis includes all samples from TCGA and non-TCGA studies with no overlapping samples, from cBioPortal ($n = 69,223$ samples). Tumor tissues are separated in two main branches

according to sample number informing the analysis (left, >3,000 samples; right, <3,000 samples). PNS, peripheral nervous system. **c**, Volcano plot showing the type of interaction for pairwise mutations in genes from the *TRIM28^{R270H}*-light hypomethylated signature in all TCGA Pan-Cancer Atlas patients. Red indicates co-occurrence, and blue indicates mutual exclusivity of pairwise mutations. One-sided Fisher’s exact test, Benjamini–Hochberg P_{adj} cutoff = 0.05. Total samples analyzed, 10,967. **d**, Our model suggests that TRIM28 buffers intrinsic developmental heterogeneity via heterochromatin silencing. By modulating a differentially methylated cancer-related gene set, it primes two distinct developmental trajectories for cancer susceptibility and outcomes, with one of the trajectories being more resistant and the other prone to cancer. Created with BioRender.com.

We also did not characterize the interaction between the *Trp53^{R270H}* and *Trim28^{D9}* alleles. While we originally expected *Trp53^{R270H/+};Trim28^{D9}* mice to exhibit the bistability phenotype driven by the *Trim28^{D9}* allele, this proved not to be the case. Although *Trp53^{R270H/+};Trim28^{D9}* double mutants exhibited increased variation in weight, the developmental bifurcation was blurred. These data suggest that *Trp53^{R270H/+}* counteracts the ability of *Trim28^{D9}* to canalize the developmental trajectories. Without confident bistability, there is no confident statistical means to separate reproducible developmental bifurcation from other roots of interindividual variation (for example, maternal–fetal interface, in utero position or litter size). p53 has important roles in obesity, metabolism and adipose tissue biology⁵⁴, and our data are consistent with double mutants exhibiting a combination of *Trim28^{D9}*-dependent (bifurcation) and *Trp53^{R270H/+}*-related obesity effects. Because TRIM28 has been shown to ubiquitinate p53 for degradation, haploinsufficiency in *Trim28* should increase the expression of p53

target genes and thus enhance DNA damage responses and counteract tumorigenesis⁵⁵. Future studies disentangling the reciprocal interaction between TRIM28 and p53 should therefore provide insight into the developmental modulation of tumorigenesis.

Finally, our data show epigenetic differences between cancer susceptibility morphs at the tenth postnatal day. Other open questions therefore are when (and where) precisely the epigenetic bifurcation takes place and whether any cell-intrinsic or -extrinsic factors might skew development toward one or the other state. The observed increased cancer incidence in male compared to female animals is interesting and mirrors human epidemiological data. The latter has been suggested to result from sex-specific differences in genetic alterations, sex chromosome-encoded genes (including epigenetic players), the immune system, hormones and metabolites⁵⁶. Our data suggest that differences in developmental priming could also be involved. Indeed, *Trim28* loss was previously associated with sex-specific defects

in embryonic epigenetic reprogramming⁵⁷, a process that appears to retain sex-specific features even in vitro⁵⁸. It is tempting to speculate that the sex-specific differences in cancer incidence observed here are due to TRIM28-mediated sex-specific developmental epigenetic effects. Our finding that TRIM28-dependent DML are enriched for a subset of human oncogenes suggests that these are embedded in a unique, more ‘permissive’ silenced state in development and hints toward a possible generalization of the model. If we (as a community) can identify similarly sensitive regions of the human cancer genome, then we will be better equipped to optimally stratify and treat patients.

Methods

Origin and maintenance of mice

This research complies with ethical regulations, with protocols approved by the Institutional Animal Care and Use Committee (Van Andel Institute (VAI); protocols 19-0026, 22-09-036, 18-10-028 and 21-08-023).

FVB/NJ. *Trim28*^{+/MommeD9} (*Trim28*^{+D9}) mice were originally generated in the Whitelaw laboratory²⁶, and B6.129S4-*Trp53*^{tm3.1Tyj/J} (*Trp53*^{+R270H}) mice were generated in the Jacks laboratory²⁸ and purchased from Jackson Laboratories (stock 008182). Both lines were backcrossed for over ten generations (FVB/NJ and C57Bl/6J, respectively) and maintained in house by breeding with wild-type siblings and periodic background refreshment using WT from JAX. Approximately 349 F₁ hybrids were generated by crossing an 8-week-old FVB. *Trim28*^{+D9} male with two 8-week-old B6. *Trp53*^{R270H/+} females, which were separated after checking for plugs the next morning. Mating mice were randomly selected. All mice were fed breeder chow (LabDiet, 5021, 0006540) ad libitum and housed in individually ventilated cages (Tecniplast, Sealsafe Plus GM500 in DGM Racks) at a maximum density of five mice per cage. Each cage was enriched with Enviro-dri (the Andersons, Crink-I’Nest) and cardboard dome homes (Shepherd, Shepherd Shack Dome). Whenever possible, same-sex siblings and same-sex animals from different litters were combined (~20 days of age) to co-house isogenic animals. Animals were kept on a 12-h light–dark cycle at an average ambient temperature of 23 °C and 35% humidity.

A total of 270 mice were randomly selected for body composition data analysis, including 137 males (18 WT, 34 *Trim28*^{+D9}, 44 *Trp53*^{+R270H} and 41 *Trp53*^{R270H/+}; *Trim28*^{+D9}) and 133 females (30 WT, 32 *Trim28*^{+D9}, 36 *Trp53*^{+R270H} and 35 *Trp53*^{R270H/+}; *Trim28*^{+D9}). At 4, 8, 16, 32, 40, 50, 60 and 70 weeks of age (or at euthanasia), mice were weighed and scanned with the EchoMRI system for fat and lean mass composition in the morning (EchoMRI, EchoMRI-100H).

Tumor analysis was conducted on 114 mice: 60 males (6 WT, 15 *Trim28*^{+D9}, 17 *Trp53*^{+R270H} and 22 *Trp53*^{R270H/+}; *Trim28*^{+D9}) and 54 females (8 WT, 8 *Trim28*^{+D9}, 21 *Trp53*^{+R270H} and 17 *Trp53*^{R270H/+}; *Trim28*^{+D9}). We performed tumor analysis blinded for genotype and phenotype, temporally collecting mice according to the timing of health reports. We specify in the text every time we are only referring to one of the sexes.

The VAI Vivarium Core officially requested a reduction of mouse cages to decrease staff in the building during the COVID-19 pandemic. As a result, we reduced active experimental mouse cages due to the extensive mouse number of our experimental cohort. Those mice are appropriately statistically censored in our data.

Mice are checked daily by animal keepers and two to three times per week by expert VAI Vivarium Core Staff for health, well-being and mass or tumor presence. Mice were flagged in health check reports if they exhibited >20% weight loss, tumors >15% of body weight as assessed by palpation (this maximal tumor size was never exceeded), tumor ulcerations, tumor discharge or hemorrhage, mobility issues, reduced appetite or hydration, limited defecation or urination, abnormal gait or posture, labored breathing, lack of movement or hypothermia. Mice with reported health concerns or those reaching the 70-week study endpoint were euthanized via CO₂ asphyxiation and cervical dislocation.

Mice used for different analysis are reported in Supplementary Table 1. The complete data of the mouse cohort are reported in Supplementary Table 2.

Genotyping

Ear punch biopsies were collected at 10 days and digested in 20 μl genomic DNA lysis buffer (100 mM Tris-HCl, pH 8.5, 5 mM EDTA, 0.2% SDS, 100 mM NaCl) with 20 mg proteinase K (Thermo Scientific, EO0491). The thermal cycling protocol used was 55 °C for 16 h, 95 °C for 10 min and a hold at 4 °C (lid at 105 °C). Nuclease-free water (Invitrogen, AM9938) was added to each lysate for a final volume of 180 μl. PCR reactions for *Trim28* and *Trp53* alleles used 1 μl diluted biopsy lysate in a 19-μl master mix (1× DreamTaq Buffer, 0.2 mM dNTPs, 0.1 μM forward and reverse primer mix, 2 U DreamTaq DNA Polymerase in nuclease-free water; Thermo Scientific, EP0703). PCR primers and thermal cycling conditions are detailed in Supplementary Tables 3 and 4. Each PCR product (20 μl) was digested with either 0.5 μl XclI and NspI (for *Trim28*^{+D9}; Thermo Scientific, FD1474) or 0.5 μl SmaI (for *Trp53*^{R270H/+}; New England Biolabs, R0571L) in a final reaction volume of 30 μl. Restriction conditions are detailed in Supplementary Table 5. Digestion products (~700 bp, WT *Trim28*; ~250 bp + ~450 bp, *Trim28*^{+D9}; ~500 bp, WT *Trp53*; ~200 bp + ~300 bp, *Trp53*^{R270H/+}) were visualized on a 3% agarose gel (Fisher Scientific, BP160-500) in 1× TAE, with GelRed as the intercalating dye (Biotium, 41003).

Statistical analyses of developmental heterogeneity

Levene’s test was used to assess homoscedasticity of body, fat or lean mass across genotypes, with *P* values adjusted by the Benjamini–Hochberg method. mclust (version 5.4.9)⁵⁹ was used for iterative expectation–maximization maximum-likelihood estimation in parameterized Gaussian mixture models, with regularization to smooth BIC. Classification uncertainty was used as the graphical parameter for fat–lean mass plots and for weighing log-rank *P* values in mouse Kaplan–Meier plots. Because most mice were classified with high confidence, the effect of this correction is negligible. Rmixmod (version 2.1.8)⁶⁰ validated mclust results via unsupervised classification and density estimation using BIC, ICL and NEC. Both methods clustered 16-week fat and lean mass data by genotype. For tests requiring it, normality and equal variances were formally tested. Data analysis was not blinded. The VAI Bioinformatics and Biostatistics Core applied generalized additive models (GAMs, version 1.22.2) to model fat and lean mass changes over time, using random-effect splines for individual slopes and intercepts. The ‘emmeans’ package (version 1.10.0)⁶¹ compared overall group fat–lean mass slope differences by group, while a separate GAM modeled fat–lean mass differences at each time point. We included a random-effect spline for each mouse but excluded the spine for a random slope by week. Cancer death proportions were analyzed with similar models, with *P* values adjusted by the Benjamini–Hochberg method.

Tissue collection

Tissues were dissected and fixed in 10% NBF solution (3.7–4% formaldehyde (37–40%), 0.03 M NaH₂PO₄, 0.05 M Na₂HPO₄ in distilled water with a final pH of 7.2 ± 0.5): epididymal white adipose tissue; uterus or preputial glands, seminal vesicles and testis; bladder; pancreas; spleen; intestine; stomach; mesenteric fat; liver; kidneys; heart; lungs; thymus; brain; breast (ninth); hindlimb muscles and bones. We also recovered spine, ribs, skull, skin and any other abnormal mass. The fixative volume was 15–20 times the tissue volume. Specimens >2.5 mm thick were cut to proper fixation. Most tissues were fixed for 40 h, while fat-rich tissues (epididymal white adipose tissue, mesenteric fat, uterus) were fixed for 72 h. Bones and spines were fixed for 1 week followed by 1 week of decalcification in 14% EDTA (14% free-acid EDTA at pH 7.2, adjusted with NH₄OH). After incubation, all tissues were transferred to 70% ethanol. Data collection was performed blinded.

Tissue preparation for histology

All tissues were embedded in paraffin by the VAI Pathology and Biorepository Core. Dehydration and clearing were automated with a Tissue-Tek VIP 5 instrument (Sakura) using the following protocol: 60 min in 70% ethanol, 60 min in 80% ethanol, 2 × 60 min in 95% ethanol, 3 × 60 min in 100% ethanol, 2 × 30 min in xylene and 75 min in paraffin. Embedding was performed with a Leica EG1150 system. Three 5- μ m sections, spaced 150 μ m apart, were cut from each tissue for hematoxylin and eosin (H&E) staining using a Leica rotary microtome. The remaining tissue was stored as a paraffin block. H&E staining was performed with a Tissue-Tek Prisma Plus Automated Slide Stainer (Sakura) and Prisma H&E Staining Kit 1.

Pathology evaluation

Standard 5- μ m H&E-stained sections were assessed for tumors and dysplastic lesions by a board-certified pathologist at the VAI Pathology and Biorepository Core. Most samples were provided blindly. Tumors were classified as malignant or benign, with all malignant tumors being primary. Metastatic or secondary tumors were identified based on primary tumor characteristics and immunohistochemical validation but were not reported in this study. Tumors were categorized into carcinomas, germ cell tumors, leukemias, lymphomas and sarcomas, with detailed classification by tissue of origin.

Tissue preparation for DNA and RNA extraction

Samples were randomly selected based on tumor type, genotype and phenotype. Curls from formalin-fixed, paraffin-embedded (FFPE) healthy tissues were prepared by the VAI Pathology and Biorepository Core, cutting three 50- μ m curls using a microtome for storage at -80 °C. A 5- μ m section was also cut for H&E staining to confirm the absence of tumors.

For tumor-containing FFPE tissues, microdissection or macrodissection was performed to separate the tumor from adjacent healthy tissue. All instrumentation and tools were treated with RNaseZAP (Invitrogen, AM9782) for decontamination, and RNase-free water was used in the tissue floating bath (Growcells.com, UPW-2000). Laser capture microdissection (LMD) was performed with the Leica LMD 6500 system. Glass slides (Leica, 11505189) were treated with UV light (Instrumedics UV curing lamp) for 30 min to prevent static and promote adherence. Mounted sections were cut at 10 μ m and dried in an oven at 60 °C for 20 min. Slides were deparaffinized with three changes of xylene for 20 min followed by one change of 100% ethanol for 20 min. All slides were dissected within 24 h of sectioning. LMD-dissected areas were stored at -80 °C until further processing. Macrodissection of FFPE tissue sections was performed manually using a razor. The mounted sections were cut at 10 μ m and dried in an oven at 60 °C for 20 min. All slides were dissected within 24 h of sectioning. Dissected areas were stored at -80 °C until further processing.

DNA extraction

Ear punch biopsies were collected and randomly selected based on tumor type, genotype and phenotype. DNA was purified using a DNeasy Blood & Tissue Kit (Qiagen, 69504) with minor modifications. After digestion, samples were brought to 220 μ l with 1 × PBS, and then steps 2–7 of the Quick-Start Protocol were followed. DNA was eluted with two washes of 100 μ l buffer AE. For samples requiring both RNA and DNA, the Quick-DNA/RNA Microprep kit (Zymo, D7005) was used, following its specific protocol.

DNA from FFPE healthy and tumor tissues was extracted with the Quick-DNA/RNA FFPE kit (Zymo, R1009), with slight modifications based on tissue source. Curls and macrodissected tissues were deparaffinized in 800 μ l Deparaffinization Solution at 55 °C for 5 min and then digested at 55 °C for 4 h. Microdissected tissues, previously deparaffinized, were digested at 55 °C for 1 h. Subsequent protocol steps were followed according to the kit manual, with an additional centrifugation step to remove residual ethanol before elution, which

used 30 μ l for macrodissected and microdissected tissues and 50 μ l for curls, respectively. Purified DNA was quantified by Qubit fluorometry (Life Technologies).

Mouse DNA methylation array

DNA samples (6–500 ng) were bisulfite converted using the Zymo EZ DNA Methylation Kit (Zymo Research) following the manufacturer's protocol with modifications for the Illumina Infinium methylation assays. After conversion, reactions were cleaned with Zymo-Spin columns and eluted in 12 μ l Tris buffer. The bisulfite-converted DNA was processed using the Illumina mouse methylation array protocol. For the assay, 7 μ l of converted DNA was denatured with 4 μ l 0.1 M NaOH. The DNA was then amplified and hybridized to the Infinium BeadChip³¹ and underwent an extension reaction with fluorophore-labeled nucleotides according to the protocol. Arrays were scanned on the Illumina iScan platform, and probe-specific calls were made using Illumina GenomeStudio version 2011.1 to generate IDAT files. Data collection and analysis were conducted blind to experimental conditions.

DNA methylation analysis

Analysis of IDAT files was performed using the SeSAmE pipeline (version 1.22.2)⁶² and its wrapper SeSAmEStr (version 0.1.0)⁶³. Fifty-eight independent biological replicates from ear biopsies of wild-type, *Trim28*^{+/D9}, *Trp53*^{R270H/+} and *Trp53*^{R270H/+}; *Trim28*^{+/D9} male mice at day 10 were analyzed. A second cohort of 22 samples from wild-type and *Trim28*^{+/D9} male mice was also analyzed independently. Additionally, 46 samples from intestines, lungs, stomachs and pancreata of *Trim28*^{+/D9} males were analyzed, comparing light and heavy mice. Data preprocessing and quality controls followed SeSAmE's default parameters and the preprocessing code 'TQCDPB', with all samples showing a detection rate >90% and no dye bias. In differential DNA methylation analyses between genotypes, the effect size cutoff was set to 0.1 (10% differential DNA methylation) with a *P*-value cutoff of <0.05, unless specified otherwise. For analyses between isogenic *Trim28*^{+/D9}-heavy and *Trim28*^{+/D9}-light mice, the effect size cutoff was 0.05 (5% differential DNA methylation), with the same *P*-value threshold. Batch effect was included as a covariate, while other technical and biological effects (detection rate, initial DNA concentration, litter) were evaluated but not included due to co-linearity with the batch effect. Global DNA methylation correlation was assessed using the 'chart.Correlation' function from the 'PerformanceAnalytics' R package (version 2.0.4). Similarity between samples was calculated as the sum of squared residuals from linear regressions. Principal-component analysis of β values was performed on SeSAmEStr pipeline output using the R function 'prcomp' from the 'stats' package (version 3.6.2). Heatmap visualization of DML was created using the R package 'ComplexHeatmap' (version 2.20.0)⁶⁴, with β values modeled and weighted using the mclust certainty score in limma (version 3.60.4)⁶⁵. Probe enrichment analysis was performed using the SeSAmE knowYourCG module, with annotations based on the knowYourCG tool. Related information is available at <http://zwdzwd.github.io/InfiniumAnnotation#mouse>. The ChromHMM annotation is derived from a mouse consensus by the ENCODE project profiling 66 mouse epigenomes across 12 tissues at daily intervals from embryonic day 11.5 to birth⁶⁶. Lists of probe-enriched genes are reported in Supplementary Table 6. Gene ontology analysis of probe-enriched genes was performed using the R package 'clusterProfiler' (version 4.12.1)⁶⁷. Probe enrichment analysis on metabolism-related gene sets was performed by retrieving all gene sets with the term 'metabolism' or 'metabolic' from the Molecular Signatures Database (MSigDB)⁶⁸. We analyzed 674 gene sets related to metabolism or metabolic processes from all MSigDB collections. Gene enrichment in the Jensen DISEASES database⁶⁹ was performed using the R package 'enrichR' (version 3.2)⁷⁰. Further data visualization of SeSAmE and SeSAmEStr output was performed using RStudio. For genomic snapshots of genes enriched for DML between *Trim28*^{+/D9}-light and *Trim28*^{+/D9}-heavy mice, we applied RUVSeq (version

1.38.0)⁷¹ to remove unwanted variation, defining empirical control probes based on differential analysis between *Trim28*^{D9}-light and *Trim28*^{D9}-heavy mice with limma⁶⁵. The 100,000 least differentially methylated probes were used in the ‘RUVg’ command from RUVSeq, and RUVSeq-corrected, z-scored β values were plotted. Data analysis was not conducted blind to the experimental conditions.

Whole-exome sequencing

Samples were randomly selected based on tumor type and genotype–phenotype combinations relevant to the biological question. Libraries for whole-exome sequencing (WES) were prepared by the VAI Genomics Core from 100–200 ng of genomic DNA using the Twist Library Preparation EF Kit 2.0 version 5.0 (Twist Bioscience). DNA was enzymatically sheared to an average size of 250 bp, followed by end repair, A tailing and ligation to uniquely barcoded dual indexes (Twist Bioscience). PCR amplification (ten cycles) was performed, and the Twist Mouse Exome Panel was used to capture whole-exome regions before a final PCR round (six cycles). Quality and quantity of the libraries were assessed using the Agilent TapeStation 4200 (Agilent Technologies) and the QuantiFluor dsDNA System (Promega). Sequencing (2 × 100 bp) was performed on an Illumina NovaSeq 6000 sequencer (Illumina) to an average raw depth of 50 million paired reads per library. Base calling was performed with Illumina RTA3 (version 3.4.4), and the output was demultiplexed and converted to FastQ format with Illumina bcl2fastq2 (version 2.20). Data collection and analysis were conducted blind to the experimental conditions.

Whole-exome sequencing analysis

Adaptor sequences and low-quality reads were trimmed using Trim Galore (version 0.6.0; <https://github.com/FelixKrueger/TrimGalore>)⁷². Trimmed reads were aligned to the mm10 reference genome with BWA (version 0.7.17)⁷³, and duplicates were marked using SAMBLASTER (version 0.1.26)⁷⁴. Paired healthy and tumor BAM files were analyzed with GATK Mutect2 (version 4.1.8.1)⁷⁵ to identify somatic single-nucleotide variants (SNVs), insertions and deletions (indels). BCFtools (version 1.17)⁷⁶ was used to extract SNVs and indels based on the gene annotation file provided by Twist Bioscience. SnpEff (version 5.1)⁷⁷ was used to predict the effect of each SNV and indel. Data visualization was performed using the ComplexHeatmap R package (version 2.20.0)⁶⁴. Data collection and analysis were not conducted blind to the experimental conditions.

RNA extraction

Fresh ear punch biopsies were collected and in part randomly selected based on tumor type and genotype–phenotype combination, without using the genotyping buffer. Rather, RNA was extracted using the Quick-DNA/RNA Microprep Plus kit (Zymo, D7005), which includes in-column DNase I treatment. For FFPE tissues, the Quick-DNA/RNA FFPE kit (Zymo, R1009) was used with minor modifications based on the tissue type. Curled and macrodissected tissues were deparaffinized in 800 μ l Deparaffinization Solution at 55 °C for 5 min, followed by digestion at 55 °C for 4 h. Already deparaffinized microdissected tissues were immediately digested for 1 h. The remaining steps were followed according to the kit manual, including except for an additional centrifugation step with the empty column before elution to remove residual ethanol. RNA was eluted with 50 μ l of DNase- and RNase-free water for curls and 30 μ l of DNase- and RNase-free water for macrodissected and microdissected tissues. Purified RNA was quantified by Qubit fluorometry (Life Technologies).

RNA sequencing

Libraries were prepared by the VAI Genomics Core from 142 ng of total RNA using the KAPA RNA HyperPrep Kit (Kapa Biosystems). Ribosomal RNA levels were reduced with the QIAseq FastSelect –rRNA HMR Kit (Qiagen). RNA was sheared to 300–400 bp, and cDNA fragments were ligated to IDT for Illumina TruSeq UD indexed adaptors (Illumina) before PCR amplification. Library quality and quantity were assessed

using the Agilent TapeStation 4200 (Agilent Technologies) and the QuantiFluor dsDNA System (Promega). Individually indexed libraries were pooled, and 2 × 100-bp sequencing was performed on an Illumina NovaSeq 6000 sequencer to an average depth of 50 million raw paired reads per transcriptome. Base calling was performed with Illumina RTA3 (version 3.4.4), and outputs were demultiplexed and converted to FastQ format with Illumina bcl2fastq2 (version 2.20). Data collection and analysis were performed blind to the experimental conditions.

RNA-sequencing analysis

All reads from FastQ files underwent quality control, adaptor trimming, mapping and counting using the mRNA-seq pipeline of snakePipes (version 2.7.3)⁷⁸. Mapping was performed with STAR (version 2.7.10b) on the GRCm38/mm10 mouse genome, and reads were counted with featureCounts (version 2.0.1). The resulting raw count matrices were input into DESeq2 (version 1.44.0)⁷⁹, accounting for technical (RIN, library preparation and sequencing batches, initial RNA concentration) and biological (litter, age) covariates. Variance-stabilizing transformation was applied and was used for downstream analyses and data visualization. Dimensional reduction analysis was performed using the Rtsne (version 0.17, *t*-SNE, <https://github.com/jkrijthe/Rtsne>) and uwot (version 0.2.2, UMAP, <https://CRAN.R-project.org/package=uwot>) R packages. Heatmap visualization was performed using the ComplexHeatmap R package (version 2.20.0)⁶⁴. Data analysis was not performed blind to the experimental conditions.

Immunofluorescence staining

Samples were randomly selected based on tumor type, genotype and phenotype. Paraffin sections (5 μ m) were deparaffinized and subjected to antigen retrieval using DAKO EnVision FLEX High pH antigen retrieval buffer for 20 min at 97 °C by the VAI Pathology and Biorepository Core. Slides were blocked with 2% FBS for 1 h and then incubated overnight at 4 °C with anti-TRIM28 (Thermo Fisher, MA5-35303, clone ARCO047) antibody diluted 1:50 in DAKO EnVision FLEX antibody diluent. After washing 3 × with 1 × PBS for 5 min, slides were stained with Rb-647 secondary antibody (Invitrogen, A32733) at 1:500 for 2 h at room temperature, followed by three washes with 1 × PBS for 5 min each. DAPI (Invitrogen, D21490) was applied for 10 min at room temperature, and slides were washed 3 × in DI water for 5 min each before coverslipping with Prolong Gold mounting medium (Invitrogen, P36930).

Immunofluorescence acquisition

Whole-tissue images were collected in a single plane on a Zeiss AxioScan 7 slide scanner using ZEN blue (version 3.7) by the VAI Imaging Core. DAPI and AF647-stained samples were excited by a Colibri 7 LED light source at 385 nm and 630 nm. Emission was detected with a Hamamatsu ORCA-Flash 4.0 camera and a Zeiss Plan-Apochromat ×20, 0.8-NA air objective. Resulting 14-bit images, scaled to 0.1725 × 0.1725 μ m per pixel, were compressed by JpgXr at 85%. Data collection and analysis were conducted blind to the experimental conditions.

Immunofluorescence analysis

Full-resolution pyramidal CZI images of TRIM28-stained tissue sections were analyzed using QuPath⁸⁰ (version 0.5.1) by the VAI Imaging Core. Annotations were made to contour target tissue while excluding bubbles, large folds, nontarget tissues and bright autofluorescence caused by debris. Nuclei were detected using the QuPath StarDist (version 0.4.4) plugin via dsb2018_heavy_augment.pb (<https://qupath.readthedocs.io/en/0.4/docs/deep/stardist.html>). QuPath’s Train Object Classifier was trained on ROIs selected from two samples of each tissue type and their batch-paired positive controls to classify nuclei as positive or negative for TRIM28. The classifier was trained using the random tree model and applied across images. Key measurements (for example, number of detections, percent positive and fluorescence) were exported for statistical analysis, with single-nuclear signals averaged per biological

replicate. At least three independent biological replicates per group (totaling 22,964,279 nuclei) were analyzed. TRIM28 nuclear fluorescence was corrected for known technical and biological confounders (antibody lot, litter, cause of death, age at death and tumor type) using `correctBatchEffects` from the R package ‘limma’ (ref. 65). Data collection and analysis were performed blind to the experimental conditions.

cBioPortal and TCGA Pan-Cancer Atlas data analyses

The TCGA Pan-Cancer Atlas³⁹, comprising 32 studies and 10,967 samples, was used for preliminary analyses in cBioPortal⁸¹. Outputs were revisualized in RStudio. Kaplan–Meier curves were generated for samples harboring mutations in either the *TRIM28*^{D9}-heavy or *TRIM28*^{D9}-light gene signatures, compared to those without mutations, unless otherwise noted. Statistical significance was tested using log-rank tests ($P < 0.05$). Co-occurrence or exclusivity of pairwise mutations within these gene signatures was tested by one-sided Fisher’s exact test, with Benjamini–Hochberg $P_{\text{adj}} < 0.05$. Overlap with the COSMIC Cancer Gene Census³⁸ was similarly tested. The effects of mutations in *TRIM28*^{D9}-light signature genes were assessed across samples from TCGA and non-overlapping samples from cBioPortal ($n = 69,223$ samples), with sex-specific analyses performed on 30,203 samples from 28,966 female patients and 34,276 samples from 32,192 male patients. Mutations were categorized as ‘harmful’ or ‘protective’ based on median survival ratios (months) between altered and unaltered samples (>1 , ‘harmful’; <1 , ‘protective’). To address sample bias, tissues were divided into low ($<3,000$) and high ($>3,000$) sample groups for heatmap visualization. The analysis was performed both on overall survival and disease-free survival. Gene expression and DNA methylation analyses of the *TRIM28*^{D9}-light signature were performed on the Pan-Cancer Atlas from TCGA samples for which both RNA-seq and DNA methylation (HM27 and HM450 arrays) data were available, comprising 10,013 samples from 10,005 patients (of a total of 10,967 samples from 10,953 patients present in the Atlas) and 33 cancer types. DNA methylation analysis focused on probes within ± 50 bp of TSS regions. Linear correlations between mRNA expression, DNA methylation and survival were visualized as heatmaps, with $P < 0.05$ for RNA-seq and 0.01 for DNA methylation. Data collection and analysis were not conducted blind to the experimental conditions.

Statistics and reproducibility

Power analysis was performed by the VAI Bioinformatics and Biostatistics Core using the `pwr` R package for power analysis (R version 3.5.2)⁸² to determine sample size. The goal was to assess cancer incidence differences (carcinoma, sarcoma, lymphoma and leukemia) between *Trim28*^{D9}-light and *Trim28*^{D9}-heavy mice. A two-sample test of proportions was performed using Firth logistic regression, with power set to 80%, $\alpha = 0.05$ and equal sample sizes assumed for each group.

Owing to COVID-related reductions, 118 mice were randomly excluded for tumor analysis. Additionally, 17 mice were excluded because they were found dead and too stiff to harvest. The final cohort included all animals from litters of 4–12 pups.

Experiments were randomized, and investigators were blinded to group allocation and outcome assessment whenever possible.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All DNA methylation array and RNA-seq data generated in this study were deposited to the Gene Expression Omnibus (GEO) under accession code [GSE262713](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE262713). The WES data generated in this study were deposited to the Sequence Read Archive (SRA) under accession code [PRJNA1094950](https://www.ncbi.nlm.nih.gov/sra/PRJNA1094950). The GRCm38/mm10 assembly is available at https://www.gencodegenes.org/mouse/release_M10.html. The MSigDB is available at

<http://www.gsea-msigdb.org/gsea/msigdb>. All cancer patient and TCGA data used in this study are publicly available on cBioPortal (<https://www.cbioportal.org>) and the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov>). Source data are provided with this paper.

Code availability

The SeSAmE wrapper pipeline SeSAmEStr is published online on Zenodo (<https://doi.org/10.5281/zenodo.7510575>)⁶³. No other custom code or mathematical algorithms were generated for this study. All publicly available codes and tools used to analyze the data are reported and referenced in the Methods. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

References

- Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
- Ganz, J. et al. Rates and patterns of clonal oncogenic mutations in the normal human brain. *Cancer Discov.* **12**, 172–185 (2022).
- García-Nieto, P. E., Morrison, A. J. & Fraser, H. B. The somatic mutation landscape of the human body. *Genome Biol.* **20**, 298 (2019).
- Lee-Six, H. et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532–537 (2019).
- Martincorena, I. et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
- Jassim, A., Rahrmann, E. P., Simons, B. D. & Gilbertson, R. J. Cancers make their own luck: theories of cancer origins. *Nat. Rev. Cancer* **23**, 710–724 (2023).
- Kakiuchi, N. & Ogawa, S. Clonal expansion in non-cancer tissues. *Nat. Rev. Cancer* **21**, 239–256 (2021).
- Baggiolini, A. et al. Developmental chromatin programs determine oncogenic competence in melanoma. *Science* **373**, eabc1048 (2021).
- Hinoue, T. et al. Analysis of the association between CIMP and BRAF^{V600E} in colorectal cancer by DNA methylation profiling. *PLoS ONE* **4**, e8357 (2009).
- Polak, P. et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**, 360–364 (2015).
- Parreno, V. et al. Transient loss of Polycomb components induces an epigenetic cancer fate. *Nature* **629**, 688–696 (2024).
- Castillo-Fernandez, J. E., Spector, T. D. & Bell, J. T. Epigenetics of discordant monozygotic twins: implications for disease. *Genome Med.* **6**, 60 (2014).
- de Oliveira Andrade, F. et al. Exposure to lard-based high-fat diet during fetal and lactation periods modifies breast cancer susceptibility in adulthood in rats. *J. Nutr. Biochem.* **25**, 613–622 (2014).
- Ekbom, A., Adami, H. O., Trichopoulos, D., Hsieh, C. C. & Lan, S. J. Evidence of prenatal influences on breast cancer risk. *Lancet* **340**, 1015–1018 (1992).
- Murugan, S., Zhang, C., Mojtahedzadeh, S. & Sarkar, D. K. Alcohol exposure in utero increases susceptibility to prostate tumorigenesis in rat offspring. *Alcohol. Clin. Exp. Res.* **37**, 1901–1909 (2013).
- Prins, G. S. Endocrine disruptors and prostate cancer risk. *Endocr. Relat. Cancer* **15**, 649–656 (2008).
- Angers, B., Perez, M., Menicucci, T. & Leung, C. Sources of epigenetic variation and their applications in natural populations. *Evol. Appl.* **13**, 1262–1278 (2020).
- Machin, G. Non-identical monozygotic twins, intermediate twin types, zygosity testing, and the non-random nature of monozygotic twinning: a review. *Am. J. Med. Genet. C Semin. Med. Genet.* **151C**, 110–127 (2009).

19. Youssoufian, H. & Pyeritz, R. E. Mechanisms and consequences of somatic mosaicism in humans. *Nat. Rev. Genet.* **3**, 748–758 (2002).
20. Feinberg, A. P. & Levchenko, A. Epigenetics as a mediator of plasticity in cancer. *Science* **379**, eaaw3835 (2023).
21. Herceg, Z. et al. Roadmap for investigating epigenome deregulation and environmental origins of cancer. *Int. J. Cancer* **142**, 874–882 (2018).
22. Ho, S.-M. et al. in *The Epigenome and Developmental Origins of Health and Disease* (ed. Rosenfeld, C. S.) 315–336 (Academic Press, 2016).
23. Ryan, R. F. et al. KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Krüppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. *Mol. Cell. Biol.* **19**, 4366–4378 (1999).
24. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932 (2002).
25. Czerwińska, P., Mazurek, S. & Wiznerowicz, M. The complexity of TRIM28 contribution to cancer. *J. Biomed. Sci.* **24**, 63 (2017).
26. Whitelaw, N. C. et al. Reduced levels of two modifiers of epigenetic gene silencing, Dnmt3a and Trim28, cause increased phenotypic noise. *Genome Biol.* **11**, R111 (2010).
27. Dalgaard, K. et al. *Trim28* haploinsufficiency triggers bi-stable epigenetic obesity. *Cell* **164**, 353–364 (2016).
28. Olive, K. P. et al. Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* **119**, 847–860 (2004).
29. Kaminsky, Z. A. et al. DNA methylation profiles in monozygotic and dizygotic twins. *Nat. Genet.* **41**, 240–245 (2009).
30. Marttila, S. et al. Methylation status of *VTRNA2-1/nc886* is stable across populations, monozygotic twin pairs and in majority of tissues. *Epigenomics* **14**, 1105–1124 (2022).
31. Zhou, W. et al. DNA methylation dynamics and dysregulation delineated by high-throughput profiling in the mouse. *Cell Genom.* **2**, 100144 (2022).
32. Filipczak, P. T. et al. p53-suppressed oncogene *TET1* prevents cellular aging in lung cancer. *Cancer Res.* **79**, 1758–1768 (2019).
33. Tovy, A. et al. p53 is essential for DNA methylation homeostasis in naive embryonic stem cells, and its loss promotes clonal heterogeneity. *Genes Dev.* **31**, 959–972 (2017).
34. Jeong, M. et al. Large conserved domains of low DNA methylation maintained by *Dnmt3a*. *Nat. Genet.* **46**, 17–23 (2014).
35. Shi, J. et al. The concurrence of DNA methylation and demethylation is associated with transcription regulation. *Nat. Commun.* **12**, 5285 (2021).
36. Turelli, P. et al. Interplay of *TRIM28* and DNA methylation in controlling human endogenous retroelements. *Genome Res.* **24**, 1260–1270 (2014).
37. Grissa, D., Junge, A., Oprea, T. I. & Jensen, L. J. Diseases 2.0: a weekly updated database of disease–gene associations from text mining and data integration. *Database* **2022**, baac019 (2022).
38. Sondka, Z. et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).
39. Weinstein, J. N. et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
40. Shen, H. et al. Integrated molecular characterization of testicular germ cell tumors. *Cell Rep.* **23**, 3392–3406 (2018).
41. Bertozzi, T. M. & Ferguson-Smith, A. C. Metastable epialleles and their contribution to epigenetic inheritance in mammals. *Semin. Cell Dev. Biol.* **97**, 93–105 (2020).
42. Muller, H. J. Types of visible variations induced by X-rays in *Drosophila*. *J. Genet.* **22**, 299–334 (1930).
43. Feinberg, A. P. & Irizarry, R. A. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc. Natl Acad. Sci. USA* **107**, 1757–1764 (2010).
44. Pujadas, E. & Feinberg, A. P. Regulated noise in the epigenetic landscape of development and disease. *Cell* **148**, 1123–1131 (2012).
45. Beer, S. et al. Developmental context determines latency of MYC-induced tumorigenesis. *PLoS Biol.* **2**, e332 (2004).
46. Alonso-Curbelo, D. et al. A gene–environment-induced epigenetic program initiates tumorigenesis. *Nature* **590**, 642–648 (2021).
47. Ohm, J. E. et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet.* **39**, 237–242 (2007).
48. Schlesinger, Y. et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat. Genet.* **39**, 232–236 (2007).
49. Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012).
50. Teschendorff, A. E. et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* **20**, 440–446 (2010).
51. Lorthongpanich, C. et al. Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. *Science* **341**, 1110–1112 (2013).
52. Hu, C. et al. Roles of Kruppel-associated box (KRAB)-associated co-repressor KAP1 Ser-473 phosphorylation in DNA damage response. *J. Biol. Chem.* **287**, 18937–18952 (2012).
53. Cassano, M. et al. Polyphenic trait promotes liver cancer in a model of epigenetic instability in mice. *Hepatology* **66**, 235–251 (2017).
54. Berkers, C. R., Maddocks, O. D. K., Cheung, E. C., Mor, I. & Vousden, K. H. Metabolic regulation by p53 family members. *Cell Metab.* **18**, 617–633 (2013).
55. Wang, C. et al. MDM2 interaction with nuclear corepressor KAP1 contributes to p53 inactivation. *EMBO J.* **24**, 3279–3290 (2005).
56. Haupt, S., Caramia, F., Klein, S. L., Rubin, J. B. & Haupt, Y. Sex disparities matter in cancer development and therapy. *Nat. Rev. Cancer* **21**, 393–407 (2021).
57. Sampath Kumar, A. et al. Loss of maternal *Trim28* causes male-predominant early embryonic lethality. *Genes Dev.* **31**, 12–17 (2017).
58. Arez, M. et al. Imprinting fidelity in mouse iPSCs depends on sex of donor cell and medium formulation. *Nat. Commun.* **13**, 5432 (2022).
59. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J.* **8**, 289–317 (2016).
60. Le Bret, R. et al. Rmixmod: the R package of the model-based unsupervised, supervised, and semi-supervised classification mixmod library. *J. Stat. Softw.* **67**, 1–29 (2015).
61. Lenth, R. *emmeans: Estimated Marginal Means, aka Least-Squares Means* [rvlenth.github.io/emmeans/](https://rpubs.io/emmeans/) (2024).
62. Zhou, W., Triche, T. J. Jr., Laird, P. W. & Shen, H. SeSAGE: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Res.* **46**, e123 (2018).
63. Apostle, S., Fagnocchi, L. & Pospisilik, J. A. SeSAMeStr: an automated pipeline for SeSAGE methylation array analysis (v1.0.0). *Zenodo* <https://doi.org/10.5281/zenodo.7510575> (2023).
64. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
65. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
66. van der Velde, A. et al. Annotation of chromatin states in 66 complete mouse epigenomes during development. *Commun. Biol.* **4**, 239 (2021).

67. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
68. Liberzon, A. et al. Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
69. Pletscher-Frankild, S., Pallegà, A., Tsafou, K., Binder, J. X. & Jensen, L. J. DISEASES: text mining and data integration of disease–gene associations. *Methods* **74**, 83–89 (2015).
70. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
71. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
72. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
73. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
74. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).
75. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
76. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
77. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w¹¹¹⁸*; *iso-2*; *iso-3*. *Fly* **6**, 80–92 (2012).
78. Bhardwaj, V. et al. snakePipes: facilitating flexible, scalable and integrative epigenomic analysis. *Bioinformatics* **35**, 4757–4759 (2019).
79. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
80. Bankhead, P. et al. QuPath: open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).
81. Cerami, E. et al. The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
82. Champely, S. et al. *pwr: Basic Functions for Power Analysis* cran.r-project.org/web/packages/pwr/ (2017).

Acknowledgements

We thank P. Laird, H. Shen, T. Yang, R. Jones, B. Williams, E. Lien, C. Essenburg, N. Vander Schaaf, P. Stevens, L. DeCamp, E. Levine, E. Ma, D. Lu, H. Lu, V. Molchanov, J. Endicott, V. Wegert, M. Edozie and D. Aicher for suggestions and support; as well as the MPI-IE Facilities and the VAI Vivarium (in particular, B. Eagleson, S. Bechaz, A. Rapp, R. Burdette, E. Tubbergen, E. Hamel, M. Powers and S. Greenwald; RRID:SCR_023211); Transgenics (in particular, T. Kempston and A. Guikema; RRID:SCR_022914); Pathology and Biorepository (in particular, S. Jewell, D. Rohrer, B. Berghuis, L. Turner, S. Whitford, A. Bouwman, K. Feenstra and K. Goudreau; RRID:SCR_022912); and the Bioinformatics and Biostatistics (RRID:SCR_024762), Optical Imaging (in particular, C. Esquibel, K. Gallik and L. Cohen; RRID:SCR_021968) and Genomics (in particular, M. Adams, M. Wegener and T. Avequin; RRID:SCR_022913) cores. We thank P. Laird, E. Lien and J. Jang for critical evaluation of the manuscript. This work was supported by funding from the MPG, the ERC and the VAI through internal philanthropy to J.A.P., NIH award R01HG012444 to J.A.P. and J.H.N., R01A1171984 and the Chan Zuckerberg Initiative with award DI-000000287 to T.J.T., R01DK132216 to J.A.P. and D. Schones, MeNU pilot project grants and a Human Frontier Science Program Long-Term Fellowship (LT000441/2018-L) to I. Panzeri. The

fundings had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

J.A.P. is the contact for the PERMUTE consortium. I. Panzeri and J.A.P. conceived the project. I. Panzeri, J.A.P., D.S. and T.J.T. designed the overall methodology, and I. Panzeri designed each individual experiment. I. Panzeri and M.T. maintained and performed the in vivo experiments, genotyping and most of the tissue collection. I. Panzeri performed DNA extraction for methylation and WES analysis and RNA extraction for RNA-seq. Y.L., K.S., C.-H.Y., A.B., A.D., E.D. and members of the PERMUTE group supported the in vivo experiments, data analysis and interpretation and genotyping. L.F. and S.A. developed the SeSAMeStr package and curated the data. I. Panzeri analyzed all the phenotypic data, while L.F. and S.A. performed genomic analyses. L.F. conceived and performed all analyses on mouse methylation arrays, mouse RNA-seq and human cancer datasets. G.H. performed all pathology reviews. I. Panzeri, E.W., Z.M. and T.J.T. performed statistical analyses of developmental heterogeneity. I. Panzeri, L.F., D.P.C. and J.A.P. wrote the original draft. I. Panzeri and L.F. prepared figures. I. Panzeri, L.F., D.P.C., J.A.P. and D.S. reviewed and edited the draft. J.A.P., T.J.T., D.P.C. and I. Panzeri acquired funds. J.A.P. provided resources for experiments. J.A.P., T.J.T. and I. Panzeri supervised the work.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at

<https://doi.org/10.1038/s43018-024-00900-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43018-024-00900-3>.

Correspondence and requests for materials should be addressed to Ilaria Panzeri or John Andrew Pospisilik.

Peer review information *Nature Cancer* thanks Mélanie Eckersley-Maslin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

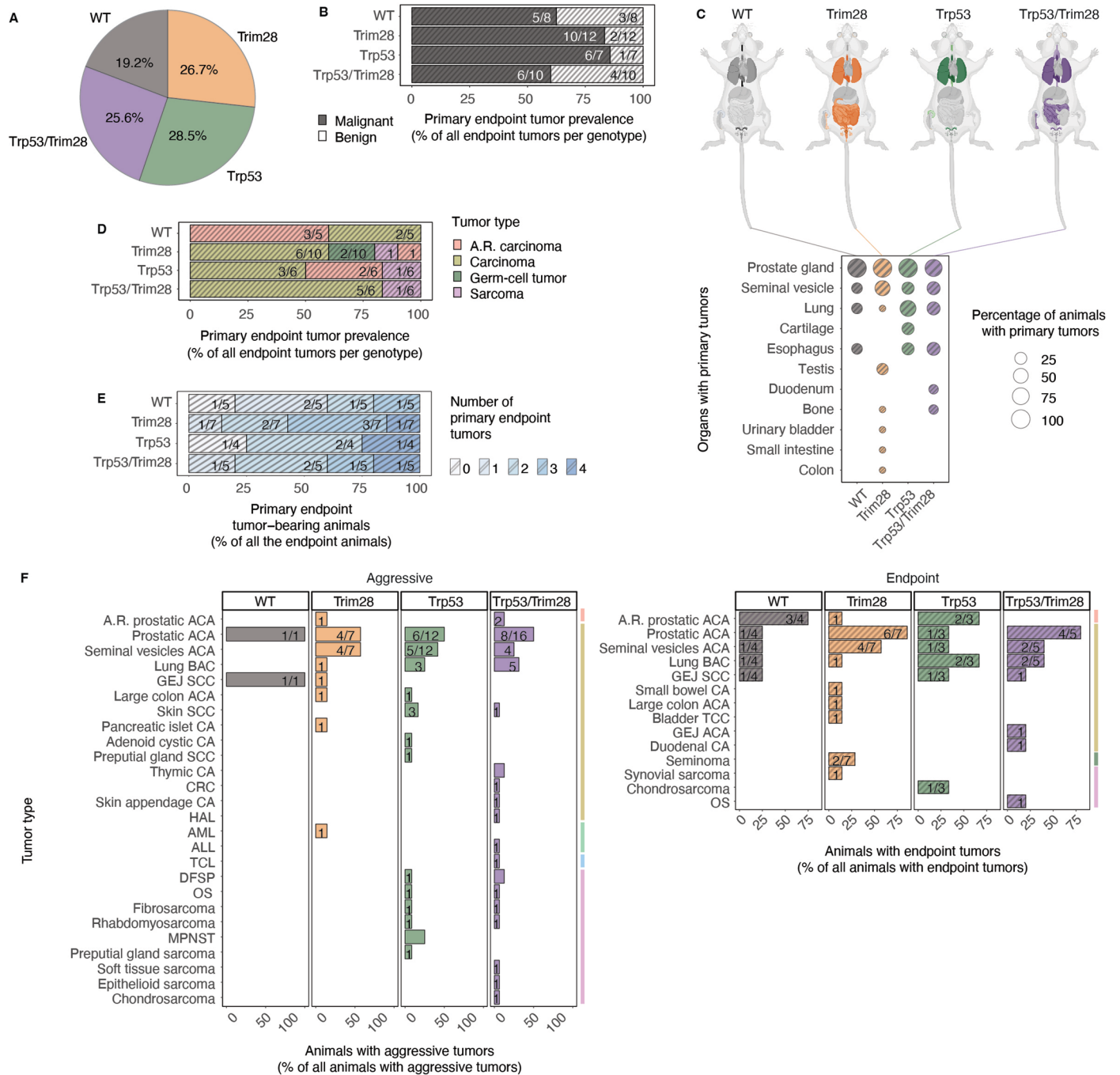
© The Author(s) 2025

¹Department of Epigenetics, Van Andel Institute, Grand Rapids, MI, USA. ²Department of Epigenetics, Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany. ³Vivarium and Transgenics Core, Van Andel Institute, Grand Rapids, MI, USA. ⁴Bioinformatics and Biostatistics Core, Van Andel Institute, Grand Rapids, MI, USA. ⁵Pathology and Biorepository Core, Van Andel Institute, Grand Rapids, MI, USA. ⁶Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, USA. ⁷Department of Biochemistry and Molecular Biotechnology, University of Massachusetts Chan Medical School, Worcester, MA, USA. ⁸Parkinson's Disease Center, Department of Neurodegenerative Science, Van Andel Institute, Grand Rapids, MI, USA. ⁹Centre for Molecular and Systems Biology, Lunenfeld-Tanenbaum Research Institute, Toronto, Ontario, Canada. ¹⁰Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. ¹¹Department of Pediatrics, MSU College of Human Medicine, East Lansing, MI, USA. ¹²Department of Translational Genomics, University of Southern California, Los Angeles, CA, USA. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: ilaria.panzeri@vai.org; andrew.pospisilik@vai.org

PERMUTE

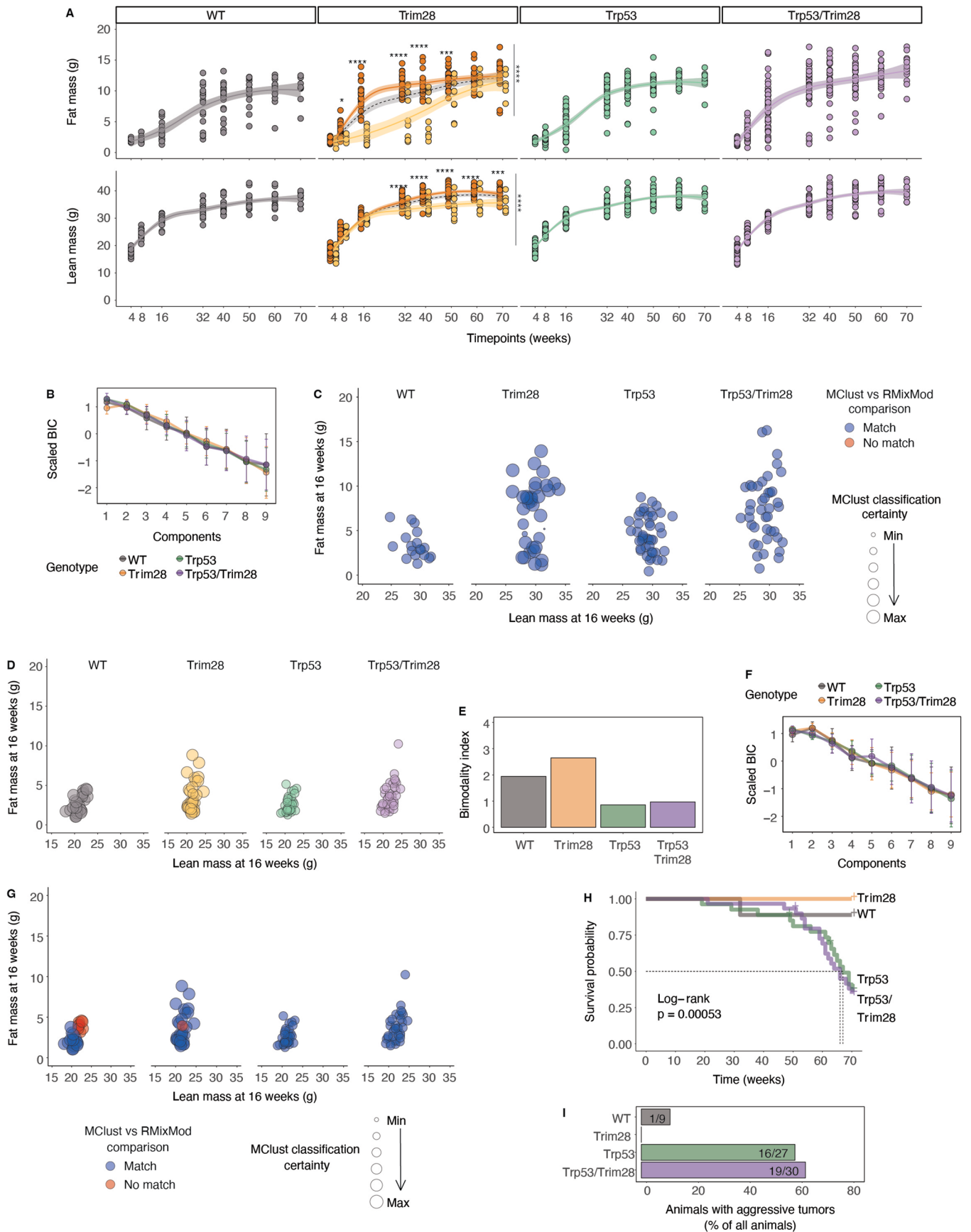
Stefanos Apostle¹, Zachary DeBruine^{1,13}, Mao Ding¹⁴, Holly Dykstra¹, Luca Fagnocchi¹, Brooke Grimaldi¹, Tim Gruber¹, Qingchu Jin¹⁵, Christine W. Lary^{14,16}, Zachary Madaj⁴, Mitchell J. McDonald¹, Joseph H. Nadeau¹⁷, Iliaria Panzeri^{1,2}, Andrea Parham¹, Ildiko Polyak¹⁷, John Andrew Pospisilik^{1,2}, Jillian Richards¹, Gabriel Seifert¹⁸, Ember Tokarski¹, Timothy J. Triche Jr.^{1,11,12}, Raimond L. Winslow¹⁵ & Emily Wolfrum⁴

¹³College of Computing, Grand Valley State University, Allendale, MI, USA. ¹⁴Department of Public Health and Health Sciences, Roux Institute at Northeastern University, Boston, MA, USA. ¹⁵Department of Bioengineering, Roux Institute at Northeastern University, Boston, MA, USA. ¹⁶Center for Interdisciplinary Population and Health Research, Maine Medical Center Research Institute, Portland, ME, USA. ¹⁷Center for Molecular Medicine, Maine Medical Center Research Institute, Portland, ME, USA. ¹⁸Department of General and Visceral Surgery, Medical Center, University of Freiburg Medical, Freiburg, Germany.



Extended Data Fig. 1 | *Trim28*^{ΔD9} male mice exhibit multi-cancer syndrome. Shading represents data from *endpoint* tumors. **a)** Distribution of genotypes in the F1 male population. N = 172 male mice. **b)** Prevalence of *malignant* (black) or *benign* (white) *endpoint* tumors by genotype, as percentage of total *endpoint* tumors per genotype. N = 37 *malignant* and *benign* *endpoint* tumors in male mice (8 in WT, 12 in *Trim28*^{ΔD9}, 7 in *Trp53*^{R270H/+}, 10 in *Trp53*^{R270H/+}; *Trim28*^{ΔD9}). **c)** Tissues targeted by *malignant* *endpoint* tumors. **Top:** mouse anatomy plots; non-targeted tissues in light-grey and targeted tissues colored by genotype: WT (black), *Trim28*^{ΔD9} (orange), *Trp53*^{R270H/+} (green), *Trp53*^{R270H/+}; *Trim28*^{ΔD9} (purple). **Bottom:** percentage of male mice with specific organs targeted by *malignant* *endpoint* tumors in the different genotypes, as percentage of total animals with *malignant* *endpoint* tumors. N = 19 mice with *malignant* *endpoint* tumors (4 WT, 7 *Trim28*^{ΔD9}, 3 *Trp53*^{R270H/+}, 5 *Trp53*^{R270H/+}; *Trim28*^{ΔD9}). **d)** Prevalence of *malignant* *endpoint* tumor types by genotype, as percentage of *malignant* *endpoint* tumors

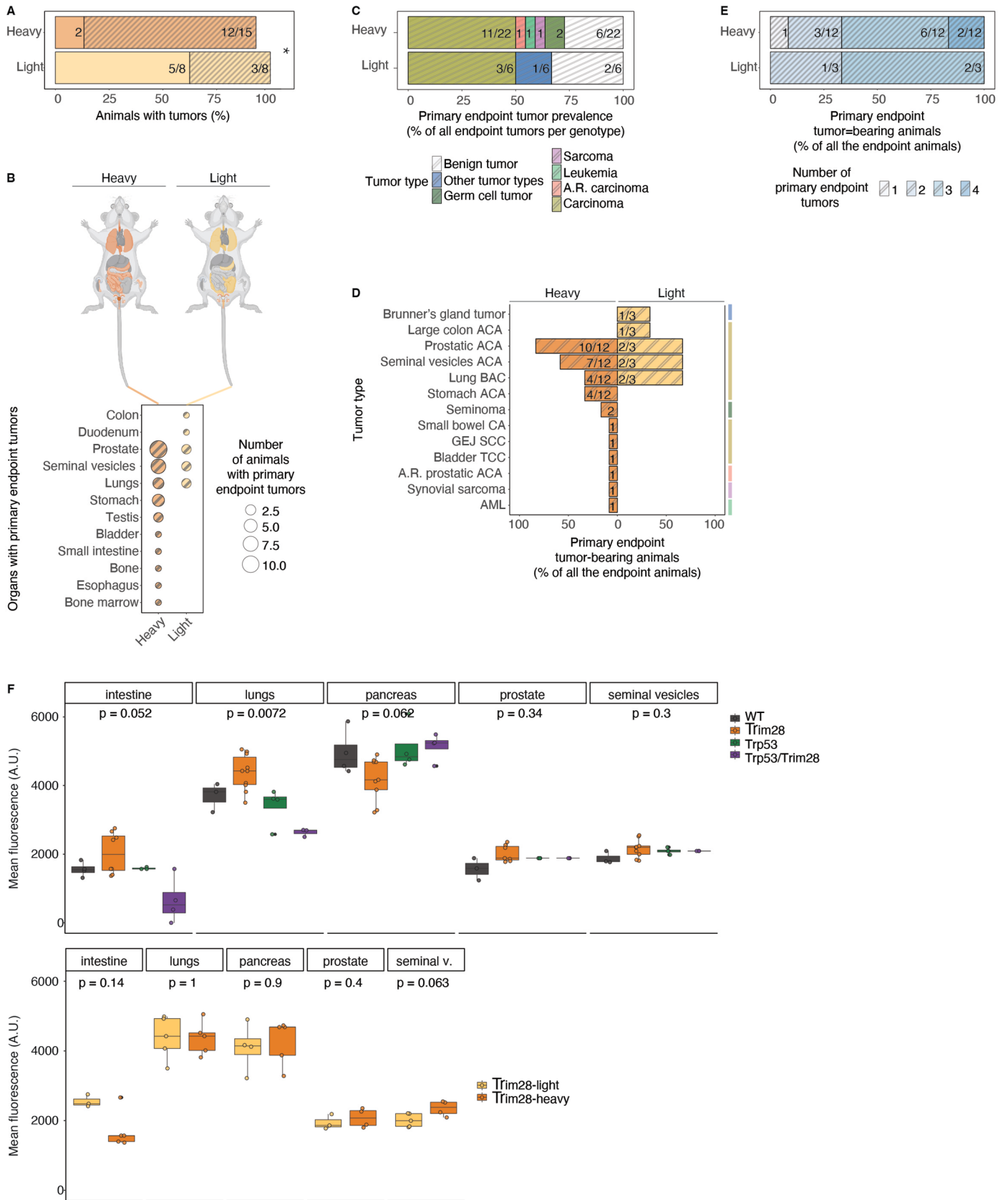
in each genotype. N = 27 total *malignant* *endpoint* tumors in male mice (5 in WT, 10 in *Trim28*^{ΔD9}, 6 in *Trp53*^{R270H/+}, 6 in *Trp53*^{R270H/+}; *Trim28*^{ΔD9}). **e)** Fraction of animals with 0 or multiple *malignant* *endpoint* tumors by genotype, as percentage of male mice screened at the *endpoint*. N = 21 *endpoint* mice (5 WT, 6 *Trim28*^{ΔD9}, 4 *Trp53*^{R270H/+}, 5 *Trp53*^{R270H/+}; *Trim28*^{ΔD9}). **f)** Prevalence of *malignant* tumor types by genotype, divided into *aggressive* (left) and *endpoint* (right), as percentage of total number male mice with *aggressive* or *endpoint* tumors, respectively. N = 36 male mice with *aggressive* tumors (1 WT, 7 *Trim28*^{ΔD9}, 12 *Trp53*^{R270H/+}, 16 *Trp53*^{R270H/+}; *Trim28*^{ΔD9}), N = 19 with *endpoint* tumors (4 WT, 7 *Trim28*^{ΔD9}, 3 *Trp53*^{R270H/+}, 5 *Trp53*^{R270H/+}; *Trim28*^{ΔD9}). Tumor categories are color-coded on the right: age-related carcinoma (red); carcinoma (gold); leukemia (light-green); lymphoma (light-blue); sarcoma (pink); germ-cell tumors (dark-green). Panel **c**, top created with **BioRender.com**.



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | TRIM28 buffers developmental heterogeneity. Each dot represents one animal mouse in all scatter plots. **a)** Scatter plots and smoothed conditional means (95% confidence interval, 'loess' method) for *fat* (**top**) and *lean* (**bottom**) mass in male mice. *Trim28^{+D9}* mice are divided into two clusters (dark- and light orange, for *-heavy* and *-light*, respectively); other genotypes have one cluster. GAM and emmeans analysis comparing *Trim28^{+D9}-heavy* vs *-light*. Significance for padj<0.0. **Top**, fat mass: p<0.0001 overall. p<0.0001 at 16, 32, 40 weeks (w); p=0.0005693 at 50w. **Bottom**, lean mass: p<0.0001 overall. p=0.0000099 at 32w, p=0.0000036 at 40w, p<0.0001 at 50w, p=0.0000085 at 60w, and p=0.0001194 at 70w. N=138 male mice (18 WT, 36 *Trim28^{+D9}*, 43 *Tp53^{R270H/+}*, 41 *Tp53^{R270H/+};Trim28^{+D9}*). **b)** Mean and standard deviation (SD) of scaled Bayesian Information Criterion (BIC) for the 10 models MClust tested for each cluster number (component) and genotype, run on same data as Fig. 2b. N=138 male mice (18 WT, 36 *Trim28^{+D9}*, 43 *Tp53^{R270H/+}*, 41 *Tp53^{R270H/+};Trim28^{+D9}*). **c)** Fat and lean mass at 16 weeks for males by genotype. Dot size proportional to MClust classification certainty, run on same data as Fig. 2b. N=138 male mice

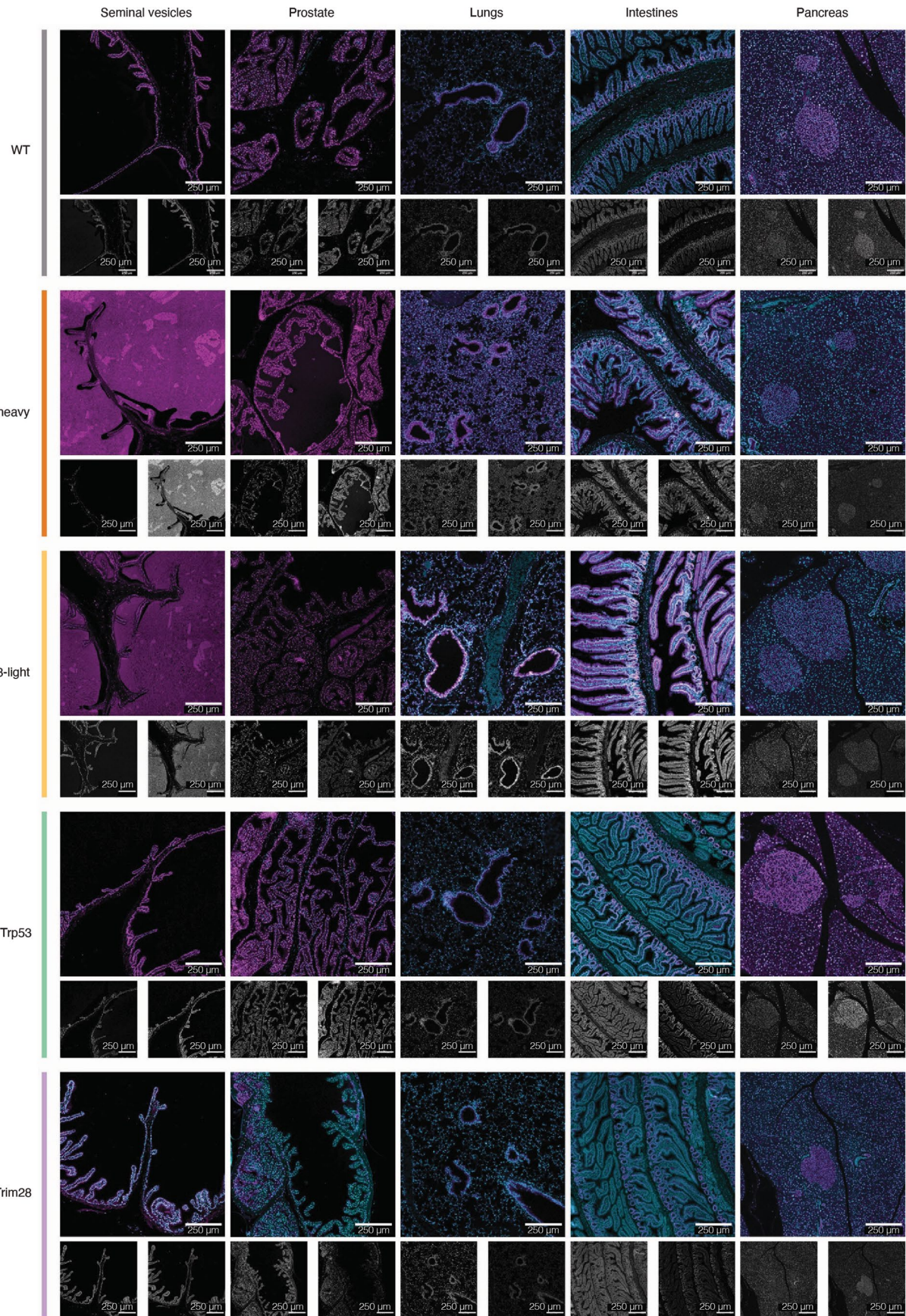
(18 WT, 36 *Trim28^{+D9}*, 43 *Tp53^{R270H/+}*, 41 *Tp53^{R270H/+};Trim28^{+D9}*). **d)** Fat and lean mass at 16 weeks for females by genotype, with overlaid density estimation by MClust. Dot size proportional to MClust classification certainty. N=133 female mice (30 WT, 32 *Trim28^{+D9}*, 36 *Tp53^{R270H/+}*, 35 *Tp53^{R270H/+};Trim28^{+D9}*). **e)** Bimodality index: ratio of MClust-determined BIC for 2 vs 1 cluster at 16 weeks, run on same data as Extended Data Fig. 3d. N=133 female mice (30 WT, 32 *Trim28^{+D9}*, 36 *Tp53^{R270H/+}*, 35 *Tp53^{R270H/+};Trim28^{+D9}*). **f)** Mean and SD of scaled BIC. N=133 female mice (30 WT, 32 *Trim28^{+D9}*, 36 *Tp53^{R270H/+}*, 35 *Tp53^{R270H/+};Trim28^{+D9}*). **g)** Fat and lean mass at 16 weeks for females by genotype. Dot size proportional to MClust classification certainty, run on same data as Extended Data Fig. 3d. N=133 female mice (30 WT, 32 *Trim28^{+D9}*, 36 *Tp53^{R270H/+}*, 35 *Tp53^{R270H/+};Trim28^{+D9}*). **h)** Kaplan-Meier survival probability by genotype. Log-rank test, p=0.00053. N=78 female mice (9 WT, 12 *Trim28^{+D9}*, 27 *Tp53^{R270H/+}*, 30 *Tp53^{R270H/+};Trim28^{+D9}*). **i)** Prevalence of female mice with *aggressive* tumors by genotype, as percentage of animals with *aggressive* (pre-endpoint) tumors over total screened animals. Actual numbers in fractions. N=78 female mice (9 WT, 12 *Trim28^{+D9}*, 27 *Tp53^{R270H/+}*, 30 *Tp53^{R270H/+};Trim28^{+D9}*).



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | TRIM28-dependent developmental heterogeneity primes cancer outcomes. Endpoint data shaded. **a**) Proportion of *Trim28^{+/D9}*-heavy and -light males with *malignant aggressive* or *endpoint* tumors, as percentage of tumor-bearing animals per group. Actual numbers in fractions. Two-sided two-sample test for equality of proportions without continuity correction, significance for $p < 0.05$: *aggressive* tumors, $p = 0.01466$, $\chi^2 = 5.9571$, $df = 1$; *endpoint* tumors, $p = 0.04153$, $\chi^2 = 4.1544$, $df = 1$. $N = 23$ male mice (15 -heavy, 8 -light). **b**) Tissues with *malignant endpoint* tumors across genotypes. **Top**: mouse anatomy plots; non-targeted tissues in light-grey and targeted tissues colored by genotype: WT (black), *Trim28^{+/D9}* (orange), *Tp53^{R270H/+}* (green), *Tp53^{R270H/+}; Trim28^{+/D9}* (purple). **Bottom**: organs with tumors in each morph. $N = 15$ male mice (12 -heavy, 3 -light). **c**) Prevalence of *endpoint* tumor types in *Trim28^{+/D9}*-heavy and -light male animals, as percentage of total *endpoint* tumors per group. Actual numbers in fractions. $N = 28$ *endpoint* tumors (*malignant* and *benign*, 22 in -heavy and 6 in -light). **d**) Distribution of *malignant endpoint* tumor types in *Trim28^{+/D9}*-heavy and -light male mice, as percentages of total *malignant endpoint* tumor-bearing animals in each group. Tumor categories are color-coded on the

right: age-related carcinoma (red); carcinoma (gold); leukemia (light-green); sarcoma (pink); germ-cell tumors (dark-green); other (teal). $N = 15$ male mice (12 -heavy, 3 -light) **e**) Fraction of *Trim28^{+/D9}*-heavy and -light males with 0 or multiple *malignant endpoint* tumors. Actual numbers in fractions. $N = 15$ male mice (12 -heavy, 3 -light). **f**) **Top**: boxplots of nuclear TRIM28 mean fluorescence intensity (MFI) across genotypes and tissues. Two-sided Kruskal-Wallis test. **Bottom**: MFI in *Trim28^{+/D9}*-heavy and -light animals across tissues. Two-sided Wilcoxon test. The lower and upper hinges of the boxplots correspond to the first and third quartiles. Boxplot hinges represent the first and third quartiles. Whiskers extend to 1.5 times the interquartile range. 22,964,279 nuclei analyzed from 38 male mice (intestine: 4 WT, 3 *Trim28^{+/D9}*-light, 5 -heavy, 3 *Tp53^{R270H/+}*, and 4 *Tp53^{R270H/+}; Trim28^{+/D9}*; lungs: 3 WT, 5 *Trim28^{+/D9}*-light, 5 -heavy, 4 *Tp53^{R270H/+}*, and 3 *Tp53^{R270H/+}; Trim28^{+/D9}*; pancreas: 4 WT, 4 *Trim28^{+/D9}*-light, 5 -heavy, 4 *Tp53^{R270H/+}*, and 4 *Tp53^{R270H/+}; Trim28^{+/D9}*; prostate: 3 WT, 3 *Trim28^{+/D9}*-light, 4 -heavy, 3 *Tp53^{R270H/+}*, and 3 *Tp53^{R270H/+}; Trim28^{+/D9}*; seminal vesicles: 3 WT, 5 *Trim28^{+/D9}*-light, 4 -heavy, 4 *Tp53^{R270H/+}*, and 3 *Tp53^{R270H/+}; Trim28^{+/D9}*). Panel **b**, top created with BioRender.com.

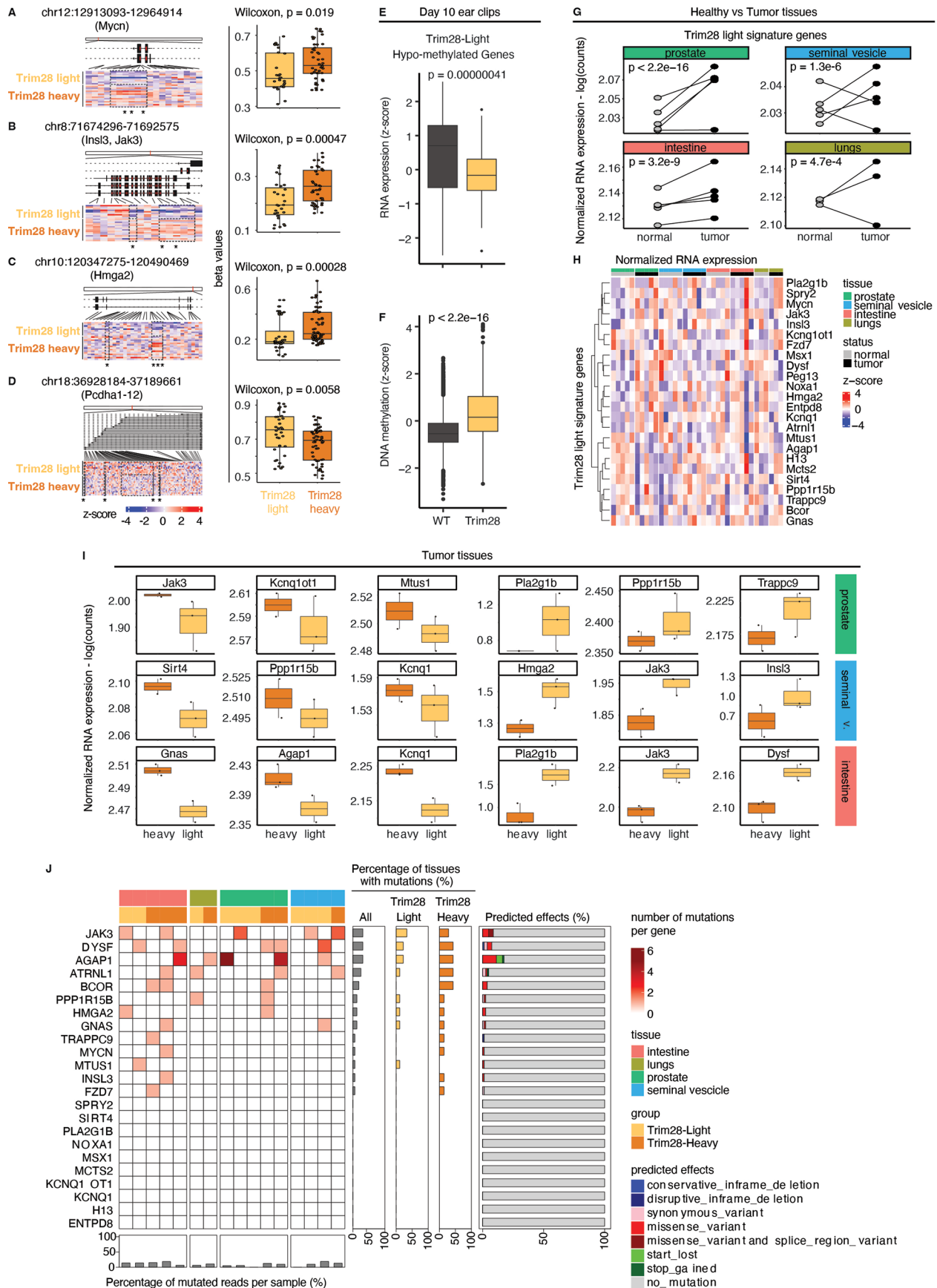


Extended Data Fig. 4 | Representative images of TRIM28 immunofluorescence in healthy tissues by genotype and in *Trim28^{+/D⁹}-heavy* and *-light* animals.

Top: analyzed tissue: seminal vesicles, prostate, lungs, intestines, and pancreas.

Left, from top to bottom: analyzed genotype: WT (light-grey), *Trim28^{+/D⁹}-*

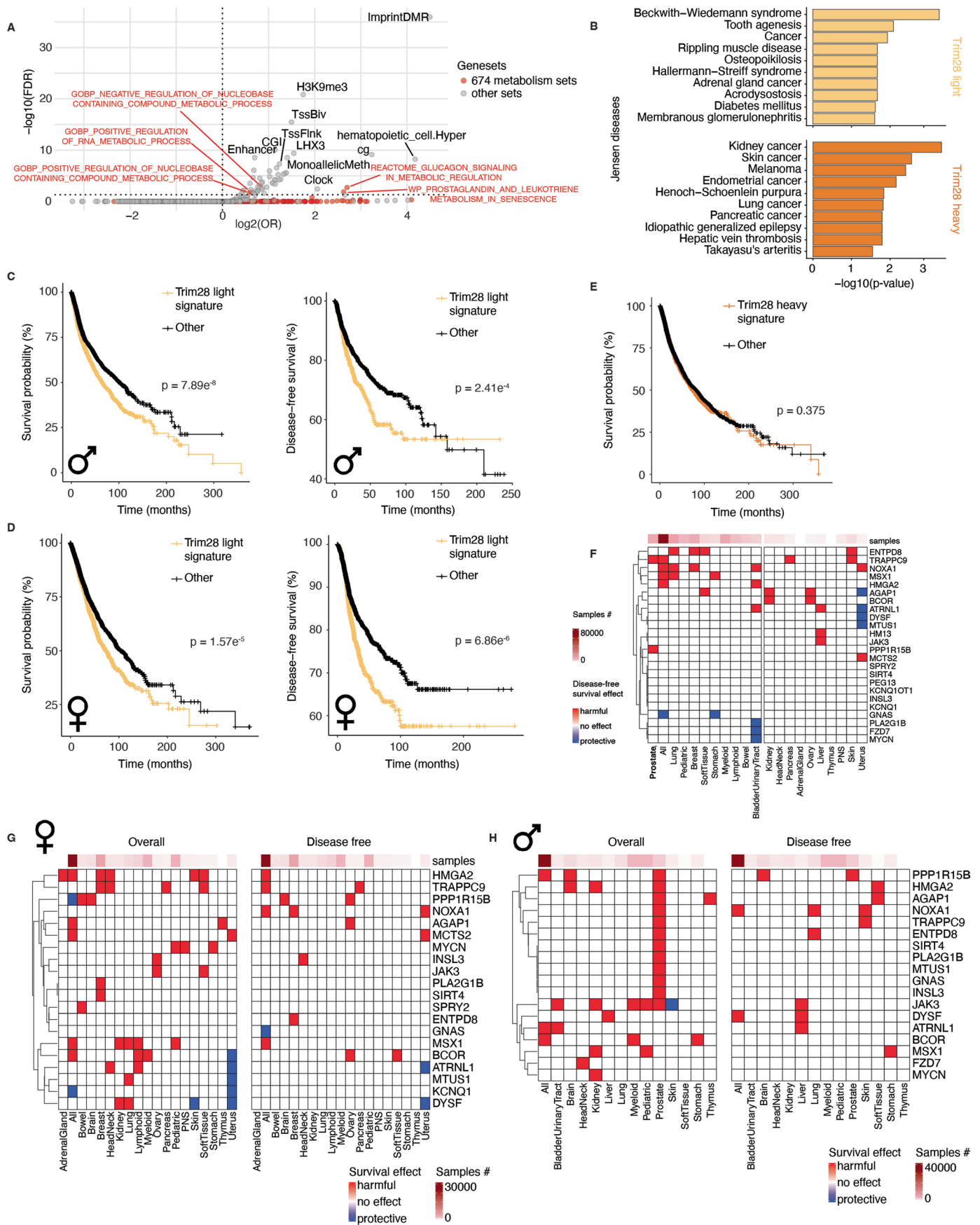
heavy (dark-orange), *Trim28^{+/D⁹}-light* (light-orange), *Trp53^{R270H/+}* (green), and *Trp53^{R270H/+}; Trim28^{+/D⁹}* (purple). Representative images from one male mouse for each tissue.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Expression profiles of *Trim28^{+/D9}-light* hypo-methylated genes in ear biopsies, healthy and tumor tissues. In all boxplots, the lower and upper hinges represent the first and third quartiles. The whiskers extend to the smallest and largest values within 1.5*the interquartile range (IQR). The solid black dot indicates the mean. **a-d**) Genomic snapshots of genes enriched for differentially methylated probes between *Trim28^{+/D9}-light* and *heavy* male mice. Black dotted boxes highlight differentially methylated regions, and significant probes are marked with asterisks. Boxplots show DNA methylation levels (beta values) for each gene. P-values are assessed by two-sided Wilcoxon test. N=24 male mice (15 *-heavy*, 9 *-light*). Genes shown: Mycn (**a**), Jak3 (**b**), Hmga2 (**c**), and the Pcdha cluster (**d**). **e**) Boxplot of mean expression for *Trim28^{+/D9}-light* hypo-methylated genes in WT vs *Trim28^{+/D9}* animals from day 10 ear biopsies. P-value = 4.1e-7, two-sided Wilcoxon test. N=15 male mice (5 WT, *Trim28^{+/D9}*). **f**) Boxplot of mean DNA methylation levels of *Trim28^{+/D9}-light* hypo-methylated probes in WT vs *Trim28^{+/D9}* from day 10 ear clips. P-value < 2.2e-16, two-sided Wilcoxon test. N=22 male mice (8 WT, 14 *Trim28^{+/D9}*). **g**) Paired dot plots showing mean

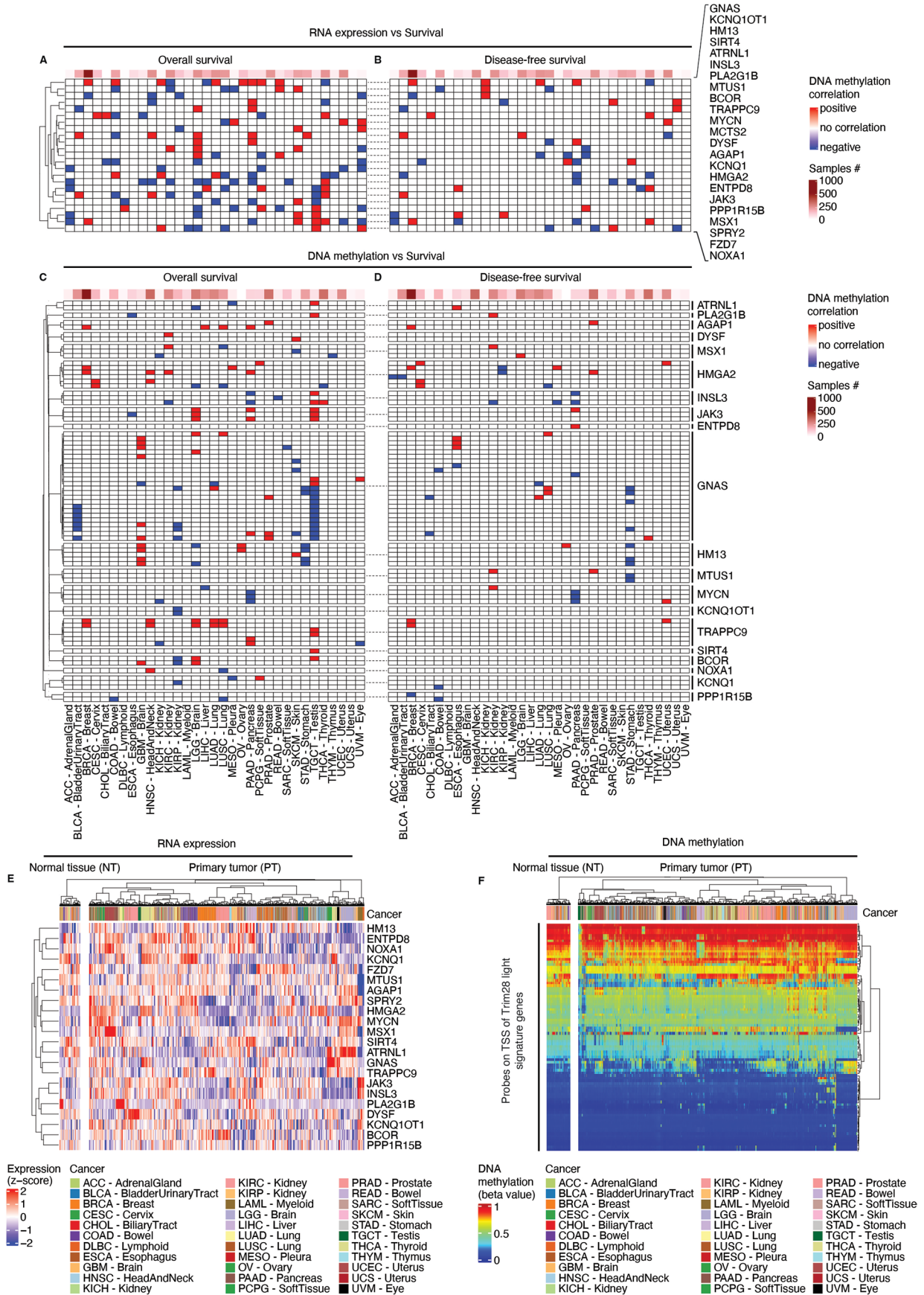
expression of *Trim28^{+/D9}-light* hypo-methylated genes in normal vs matched tumor samples across tissues. P-values: prostate (<2.2e-16), seminal vesicles (1.3e-6), intestine (3.2e-9), lungs (4.7e-4), two-sided Wilcoxon test. **h**) Heatmap of *Trim28^{+/D9}-light* hypo-methylated genes in normal and matched tumor tissues, with Z-score transformation. In **g-h**, N=36 tissues (5 healthy and 5 matched tumor biological replicates for prostate, seminal vesicle, and intestine tumors, and 3 healthy and 3 matched lung tumor biological replicates, all from *Trim28^{+/D9}* animals). **i**) Representative boxplots showing expression of *Trim28^{+/D9}-light* hypo-methylated genes in tumor samples from *Trim28^{+/D9}-light* and *-heavy* tissues. N=18 tissues (prostate: 3 *-light*, 2 *-heavy* animals; seminal vesicle: 3 *-light*, 2 *-heavy*; intestine: 2 *-light*, 3 *-heavy*; lungs: 1 *-light*, 2 *-heavy*). **j**) Heatmap showing number of mutations for indicated genes in tissues from *Trim28^{+/D9}-light* and *-heavy* animals. Relative percentages of samples with mutations, and predicted effects, are reported alongside. Under the heatmap, bar plots show the percentage of mutated reads per sample. N=16 tissues (5 intestines, 2 lungs, 5 prostate, 4 seminal vesicles).



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Sex-specific effect of the *Trim28*^{+/D9}-light hypo-methylated signature in human patients. a) Volcano plot showing enrichment of metabolism-related gene sets (red) or other gene sets (grey) on *Trim28*^{+/D9}-light hypo-methylated probes. The dotted line marks the False Discovery Rate (FDR) cut-off at 0.05. N=24 male mice (15 *Trim28*^{+/D9}-heavy, 9 -light). **b)** Enrichment of genes with hypo-methylated probes in *Trim28*^{+/D9}-light (light-orange) vs -heavy male animals (dark-orange) using the Jensen DISEASES database. P-value cut-off = 0.05, Fisher's exact test. N=24 male mice (15 *Trim28*^{+/D9}-heavy, 9 -light). **c)** Kaplan-Meier survival curves in male patients. Left: overall survival in months, showing significant difference between patients with mutations in *Trim28*^{+/D9}-light hypo-methylated signature genes (light-orange, n=1660 samples) vs. non-mutated (black, n=3175 samples). Log-rank test, p=7.89e-8. Right: disease-free survival, comparing 687 mutated vs. 1478 non-mutated patients, same color-code as in the left panel. Log-rank test, p=2.41e-4. **d)** Kaplan-Meier survival

curves in female patients. Left: overall survival comparing 1903 mutated samples (light-orange) vs. 3391 non-mutated samples (black). Log-rank test, p=1.57e-5. Right: disease-free survival, comparing 1125 mutated patients (light-orange) vs. 2023 non-mutated patients (black). Log-rank test, p=6.86e-6. **e)** Kaplan-Meier curve for overall survival of all TCGA Pan-Cancer patients with mutations in *Trim28*^{+/D9}-heavy hypo-methylated genes (dark-orange, n=1865 samples) vs. non-mutated (black, n=9085 samples), showing no significant difference (p=0.375). Total cases = 10967 patients. **f)** Heatmap of disease-free survival impacts by mutations in specified genes and tumor types from TCGA/non-TCGA samples (N=69223 samples). Tumors are divided into two branches based on sample size (>3000, <3000). **g-h)** Heatmaps showing effects of gene expression on overall and disease-free survival in female (**g**) and male (**h**) patients, with a P-value cut-off = 0.05 from two-sided Pearson's correlations. Female: 5318 patients; Male: 4871 patients.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Expression and DNA methylation of the *Trim28^{+/D9}-light* hypo-methylated signature in human patients, and their effects.

a) Heatmap of the effects on overall survival probability of expression of the indicated genes and tumor types. P-value cut-off = 0.05, by linear regression analysis. N=10013 samples. **b)** Heatmap of the effects on disease-free survival probability of expression of the indicated genes and tumor types. P-value cut-off = 0.05, by linear regression analysis. N=10013 samples. **c)** Heatmap of the effects on overall survival probability of DNA methylation at Transcription Start Sites (TSS) of the indicated genes and tumor types. P-value cut-off = 0.01, by linear regression analysis. N=10013 samples. **d)** Heatmap of the effects on disease-free survival probability of DNA methylation at TSS of the indicated genes and tumor types. P-value cut-off = 0.01, by linear regression analysis. N=10013 samples. **e)** Heatmap showing the expression profile of *Trim28^{+/D9}-light* signature genes, across tumor types from the TCGA Pan-Cancer Atlas. Primary tumors are separated from the available matching normal tissues. N=10013 samples. **f)** Heatmap showing DNA methylation profile at TSS of all *Trim28^{+/D9}-light* signature genes, across tumor types from the TCGA Pan-Cancer Atlas. Primary tumors are separated from the available matching normal tissues. N=10013 samples. In all panels, the analyses include all samples from the TCGA Pan-Cancer Atlas with

available RNA-seq and DNA methylation array data (N=10013 samples). TCGA Pan-Cancer Atlas legend: LAML=Acute Myeloid Leukemia; ACC=Adrenocortical carcinoma; BLCA=Bladder Urothelial Carcinoma; LGG=Brain Lower Grade Glioma; BRCA=Breast invasive carcinoma; CESC=Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL=Cholangiocarcinoma; LCML=Chronic Myelogenous Leukemia; COAD=Colon adenocarcinoma; ESCA=Esophageal carcinoma; GBM=Glioblastoma multiforme; HNSC=Head and Neck squamous cell carcinoma; KICH=Kidney Chromophobe; KIRC=Kidney renal clear cell carcinoma; KIRP=Kidney renal papillary cell carcinoma; LIHC=Liver hepatocellular carcinoma; LUAD=Lung adenocarcinoma; LUSC=Lung squamous cell carcinoma; DLBC=Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; MESO=Mesothelioma; MISC=Miscellaneous; OV=Ovarian serous cystadenocarcinoma; PAAD=Pancreatic adenocarcinoma; PCPG=Pheochromocytoma and Paraganglioma; PRAD=Prostate adenocarcinoma; READ=Rectum adenocarcinoma; SARC=Sarcoma; SKCM=Skin Cutaneous Melanoma; STAD=Stomach adenocarcinoma; TGCT=Testicular Germ Cell Tumors; THYM=Thymoma; THCA=Thyroid carcinoma; UCS=Uterine Carcinosarcoma; UCEC=Uterine Corpus Endometrial Carcinoma; UVM=Uveal Melanoma.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection
Illumina Genome Studio v2011.1 - DNA methylation array scanning
Illumina RTA3 v3.4.4 - RNA-seq and WES base calling
Illumina Bcl2fastq2 v2.20 - RNA-seq and WES FastQ conversion

Data analysis
Publicly available codes:
SeSAMe v1.22.2 - DNA methylation array analysis
snakePipes v2.7.3 - RNA-seq QC, mapping, and features counts
STAR v2.7.10b - RNA-seq reads mapping
featureCounts v2.0.1 - RNA-seq reads count on features
ComplexHeatmap v2.20.0 - Heatmap visualization
Limma v3.60.4 - Differential analysis
DESeq2 v1.44.0 - RNA-seq differential analysis
clusterProfiler v4.12.1 - Gene ontology analysis
Rtsne v0.17 - tSNA
uwot v0.2.2 - UMAP
RUVSeq v 1.38.0 - remove unwanted variation from RNA-Seq Data
PerformanceAnalytics v2.0.4 - Correlation heatmap
enrichR v3.2 - Gene ontology analysis
MClust v5.4.9 - Gaussian finite mixture modeling
Rmixmod v2.1.8 - Mixture modeling classification
gam - Generalized additive modeling

emmeans - Estimated Marginal Means
 stats v3.6.2 - Statistical analysis and PCA
 Trim Galore v0.6.0 - WES trimming
 bwa v0.7.17 - WES mapping
 samblaster v0.1.26 - WES duplicates marker
 GATK Mutect2 v4.1.8.1 - WES indels and SNV caller
 BCFtools v1.17 - WES indels and SNV extraction
 SnpEff v5.1 - WES SNV effect prediction
 QuPath v0.5.1 - full resolution pyramidal czi images import
 QuPath StarDist v0.4.4 - nuclei detection
 ZEN blue v3.7 - images collection

Custom codes:

SeSAmEstr v0.1.0 - DNA methylation array analysis

The SeSAmE wrapper pipeline SeSAmEstr is published online in Zenodo: <https://doi.org/10.5281/zenodo.7510575>. No other custom code or mathematical algorithms were generated for this study. All publicly available codes and tools used to analyze the data are reported and referenced in the methods sections. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All DNA methylation array and RNA sequencing data generated in this study were deposited to the Gene Expression Omnibus (GEO) under accession code GSE262713. The WES data generated in this study were deposited to the Sequence Read Archive (SRA) under accession code PRJNA1094950. The GRCm38/mm10 is available at https://www.gencodegenes.org/mouse/release_M10.html. The MSigDB is available at <http://www.gsea-msigdb.org/gsea/msigdb>. All cancer patients and TCGA data used in this study are publicly available at the cBioPortal (<https://www.cbioportal.org>) and the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov>). Source data for all Figures and Extended Data Figures have been provided as Source Data files.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Reporting on race, ethnicity, or other socially relevant groupings

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The sample size of our animal experiments was determined by power calculation. Our experiments were designed to have >95% power to detect phenotypic variation (i.e. identify bi-modal distributions) and are validated across mouse houses. Our power calculation is based on Hartigan's dip test performed on 1000 simulations of bimodal data assuming each mode is roughly Gaussian, and with each mode scaled by its

variance. A post-hoc analysis showed we have >99% power at 16 weeks of age. Sex was considered in the study design, and all the findings were reported both in male and female animals, in distinct panels.

Data exclusions	118 animals were excluded from the tumor analysis because of forced reduction in mouse numbers due to the COVID lockdown. 17 mice were not analyzed because found dead and already too stiff to be harvested. Dead mice due to dehydration were also excluded from the study. The resulting cohort included all animals coming from litters comprising 4-12 pups.
Replication	Trim28+/D9-mediated developmental bi-stability was replicated in two independent mouse houses, even after independent re-derivations from cryo-preserved embryos and sperm. For all experiments, at least 3 independent biological replicates were performed. All replication attempts were successful. All attempts at replication are included in the data.
Randomization	Mice are sex- and littermate- matched. All samples used for DNA methylation arrays, RNA-seq, WES were randomized during processing and sequencing. For human data, a sex-matched analysis of all cancer patients data was performed, with samples assigned to either the control or diseased groups.
Blinding	Body composition data were collected by the researcher without knowing the genotypes. Data analysis were performed by researchers blinded to group/phenotype information.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study	n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants		

Antibodies

Antibodies used	TRIM28 (ThermoFisher, MA5-35303, clone ARC0047) antibody Rb-647 (Invitrogen, A32733), secondary antibody
Validation	Validated against a synthetic peptide corresponding to a sequence within amino acids 250-350 of human KAP1/TRIM28 (Q13263). TRIM28 Recombinant Monoclonal antibody specifically detects TRIM28 in Human, Mouse, Rat samples. It is validated for Immunohistochemistry (Paraffin), Immunoprecipitation, Western Blot

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	FVB/NJ.Trim28+/MommeD9 (Trim28+/D9) - generated in the Whitelaw lab B6.129S4-Trp53<tm3.1Tyj>/J (Tp53+/R270H) - Jackson Laboratories (stock #008182) F1 hybrids were generated by crossing 8-week-old FVB.Trim28+/D9 males with 8-week-old B6.Tp53R270H/+ females. From these crosses, we generated FVB.Trim28+;/+;B6.Tp53+;/+, FVB.Trim28+/D9;B6.Tp53+;/+, FVB.Trim28+;/+, B6.Tp53R270H/+ and FVB.Trim28+/D9;B6.Tp53R270H/+ animals.
Wild animals	This study did not involve wild animals.
Reporting on sex	Male and female offspring were analyzed and considered in the study design; females exhibited low levels of both phenotypic heterogeneity and cancer incidence, precluding robust analysis with enough statistical power.
Field-collected samples	The study did not involve field-collected samples.
Ethics oversight	All protocols were approved by Institutional Animal Care and Use Committee under protocols 19-0026, 22-09-036, 18-10-028, and 21-08-023, at Van Andel Institute (VAI, USA). Routine health monitoring was performed 2-3 times per week and mice were euthanized anytime tumors reached a maximum size of ~15% of total body weight (or below). Maximal tumor size was never exceeded.

Note that full information on the approval of the study protocol must also be provided in the manuscript.