Aldehyde-driven transcriptional stress triggers an anorexic DNA damage response

https://doi.org/10.1038/s41586-021-04133-7

Received: 16 June 2020

Accepted: 13 October 2021

Published online: 24 November 2021

Check for updates

Lee Mulderrig[1,2,11], Juan I. Garaycoechea[3,11], Zewen K. Tuong[4,5], Christopher L. Millington[1], Felix A. Dingler[1], John R. Ferdinand[4], Liam Gaul[6], John A. Tadross[7,8], Mark J. Arends[9], Stephen O'Rahilly[7], Gerry P. Crossan[2], Menna R. Clatworthy[4,5,10] & Ketan J. Patel[1,2]✉

Endogenous DNA damage can perturb transcription, triggering a multifaceted cellular response that repairs the damage, degrades RNA polymerase II and shuts down global transcription[1–4]. This response is absent in the human disease Cockayne syndrome, which is caused by loss of the Cockayne syndrome A (CSA) or CSB proteins[5–7]. However, the source of endogenous DNA damage and how this leads to the prominent degenerative features of this disease remain unknown. Here we find that endogenous formaldehyde impedes transcription, with marked physiological consequences. Mice deficient in formaldehyde clearance (*Adh5*$^{-/-}$) and CSB (*Csb*$^{m/m}$; *Csb* is also known as *Ercc6*) develop cachexia and neurodegeneration, and succumb to kidney failure, features that resemble human Cockayne syndrome. Using single-cell RNA sequencing, we find that formaldehyde-driven transcriptional stress stimulates the expression of the anorexiogenic peptide GDF15 by a subset of kidney proximal tubule cells. Blocking this response with an anti-GDF15 antibody alleviates cachexia in *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice. Therefore, CSB provides protection to the kidney and brain against DNA damage caused by endogenous formaldehyde, while also suppressing an anorexic endocrine signal. The activation of this signal might contribute to the cachexia observed in Cockayne syndrome as well as chemotherapy-induced anorectic weight loss. A plausible evolutionary purpose for such a response is to ensure aversion to genotoxins in food.

Damage to the DNA template poses a challenge to the transcriptional machinery, stalling the progression of RNA polymerase II (Pol II)[1,2]. This stalling initiates the repair of the damaged DNA template through transcription-coupled nucleotide excision repair (TC-NER)[3] and the degradation of Pol II, thereby globally shutting down transcription[4]. The CSB protein triggers TC-NER by binding the stalled Pol II and recruiting downstream nucleotide excision repair (NER) factors to excise the damage[3,8]. If TC-NER fails, a last-resort mechanism ensures that Pol II is degraded[9].

The importance of responding to such transcriptional stress is evident from the severe phenotype of human patients with Cockayne syndrome, which is caused by mutations in *CSA* (also known as *ERCC8*) or *CSB* (also known as *ERCC6*). These patients develop progressive cachexia, neurodegeneration and kidney failure[5–7]. The identity of the endogenous factor or factors that cause the damage leading to stalling of Pol II and how this damage ultimately drives the severe phenotype of Cockayne syndrome remain unknown. Addressing this fundamental question has been challenging because mice that lack either CSA or CSB do not show features of human Cockayne syndrome[10,11]. Here we reveal that endogenous formaldehyde drives the Cockayne syndrome phenotype in mice, causing the kidney to express the anorectic factor GDF15, which stimulates weight loss.

Formaldehyde causes transcriptional stress

Formaldehyde was previously identified as an important source of endogenous DNA damage, necessitating DNA interstrand crosslink repair[12]. This aldehyde is highly reactive, and is capable of causing a variety of DNA lesions[13]. We set out to test whether NER is required to repair lesions caused by formaldehyde. NER has two distinct branches: global genome NER (GG-NER), which removes DNA lesions throughout the genome by sensing distortions of the DNA helix; and TC-NER, which repairs damage in transcribed DNA[14]. We generated transformed mouse embryonic fibroblast (tMEF) lines that lack GG-NER (*Xpc*$^{-/-}$), TC-NER (*Csb*$^{m/m}$), or both GG-NER and TC-NER (*Xpa*$^{-/-}$), alone or in combination with the formaldehyde-detoxifying enzyme ADH5. We found that *Xpa*$^{-/-}$ and *Csb*$^{m/m}$ cells were hypersensitive to formaldehyde (Fig. 1a, b), in contrast to *Xpc*$^{-/-}$ cells (Fig. 1c). Furthermore, combined ablation of *Adh5* conferred additional formaldehyde sensitivity when either *Xpa* or *Csb* were inactivated (Fig. 1a, b) but not when *Xpc* was inactivated (Fig. 1c), indicating that both formaldehyde detoxification and TC-NER (but not GG-NER) are required to protect cells against formaldehyde toxicity. CSB is also involved in the repair of oxidative damage

[1]MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK. [2]MRC Laboratory of Molecular Biology, Cambridge, UK. [3]Hubrecht Institute, Utrecht, The Netherlands. [4]Molecular Immunity Unit, Department of Medicine, University of Cambridge, Cambridge, UK. [5]Cellular Genetics, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. [6]The Francis Crick Institute, London, UK. [7]MRC Metabolic Diseases Unit, Wellcome Trust-Medical Research Council Institute of Metabolic Science, University of Cambridge, Cambridge, UK. [8]Department of Pathology, University of Cambridge, Cambridge, UK. [9]Division of Pathology, University of Edinburgh, Cancer Research UK Edinburgh Centre, IGMM, Edinburgh, UK. [10]Cambridge Institute of Therapeutic Immunology and Infectious Diseases, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Cambridge, UK. [11]These authors contributed equally: Lee Mulderrig, Juan I. Garaycoechea. ✉e-mail: ketan.patel@imm.ox.ac.uk

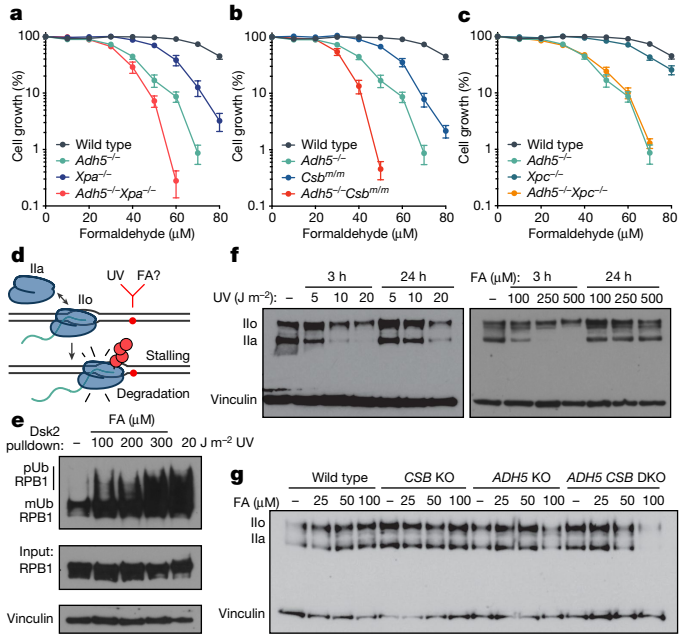


Fig 1 | Formaldehyde causes transcriptional stress. a–c, Cytotoxicity of formaldehyde in *Xpa*$^{-/-}$ and *Adh5*$^{-/-}$*Xpa*$^{-/-}$ (**a**), *Csb*$^{m/m}$ and *Adh5*$^{-/-}$*Csb*$^{m/m}$ (**b**), and *Xpc*$^{-/-}$ and *Adh5*$^{-/-}$*Xpc*$^{-/-}$ (**c**) tMEF cell lines (triplicate experiments; data are mean ± s.e.m.). **d**, Scheme outlining the poly-ubiquitylation and degradation of Pol II in response to UV and formaldehyde (FA). **e**, Dsk2 pulldown of mono-ubiquitylated (mUb) and poly-ubiquitylated (pUb) RPB1 in HEK 293 cells 30 min after treatment with formaldehyde. **f**, RPB1 degradation in HEK 293 cells after UV and formaldehyde treatment. **g**, RPB1 degradation in HEK 293 cells following 24 h of formaldehyde treatment. KO, knockout; DKO, double-knockout. Uncropped western blots are presented in Supplementary Fig. 1.

independently of its role in TC-NER[15]. We complemented cells with a separation-of-function point mutant of CSB (K991R), which has intact TC-NER but is defective in oxidative damage repair. We observed that the K991R mutant showed no increase in sensitivity to formaldehyde, similar to UV exposure (Extended Data Fig. 1a–c). These data indicate that it is likely to be the role of CSB in TC-NER and not in oxidative damage repair that contributes to cellular protection against formaldehyde.

Actively transcribing Pol II (IIo) stalls when it encounters transcription-blocking lesions caused by UV. The largest Pol II subunit (RPB1) is then rapidly poly-ubiquitylated and degraded. After degradation of IIo, the non-transcribing Pol II (IIa) then begins transcribing, becoming IIo and depleting the pools of IIa[4,9] (Fig. 1d). To determine whether formaldehyde elicits the same response, we first confirmed our tMEF observations in HEK 293 cells (Extended Data Fig. 1d, e). Then, we exposed wild-type HEK 293 cells to UV or formaldehyde and found that formaldehyde exposure stimulates the poly-ubiquitylation and degradation of Pol II in a dose-dependent manner (Fig. 1e, f). Finally, we tested the stability of Pol II in *ADH5* and *CSB* single- and double-knockout cells after exposure to formaldehyde (Fig. 1g). This revealed that cells lacking both ADH5 and CSB are more sensitive to formaldehyde-induced Pol II degradation compared with controls. In summary, both detoxification and CSB protect cells against formaldehyde toxicity, and formaldehyde exposure leads to transcription-blocking DNA damage that triggers Pol II poly-ubiquitylation and degradation.

Formaldehyde drives Cockayne syndrome

To test whether sufficient formaldehyde is generated endogenously to necessitate repair by NER, we crossed mice that lacked ADH5 with mice lacking genes involved in NER (*Xpa*, *Xpc* or *Csb*). In agreement with previous data, NER$^{+/-}$ inter-crosses yielded NER$^{-/-}$ mice at expected Mendelian ratios (Extended Data Fig. 2a). However, in the absence of ADH5, *Adh5*$^{-/-}$*Xpa*$^{-/-}$ and *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice were significantly underrepresented (7.5% and 5.4%, respectively, instead of 25%), whereas *Adh5*$^{-/-}$*Xpc*$^{-/-}$ mice were born at the expected ratio (Extended Data Fig. 2b). Therefore, in agreement with our cell line data, *Adh5* genetically interacts with genes required for TC-NER but not with genes required for GG-NER. Additionally, we noted that if the mothers were *Adh5*$^{+/-}$, the frequency of *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice born was partially rescued (Extended Data Fig. 2c), indicating that maternal formaldehyde clearance has a role in preserving the development of *Adh5*$^{-/-}$*Csb*$^{m/m}$ embryos.

Adh5$^{-/-}$*Csb*$^{m/m}$ mice were smaller in size and showed a marked age-dependent reduction in fat. Notably, both features were much more prominent in males than in females (Fig. 2a, Extended Data Fig. 2e, 3a). Mass reduction was not observed in *Adh5*$^{-/-}$*Xpc*$^{-/-}$ mice or in *Adh5*$^{-/-}$*Xpa*$^{-/-}$ males; these two lines appeared phenotypically normal (Extended Data Fig. 2d). Aged *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice developed neurological deficits including ataxia, hind limb clasping (Fig. 2b) and kyphosis, along with a decline in grip-strength performance (Extended Data Fig. 3b, c). Brains of *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice became significantly smaller with age (Extended Data Fig. 3d) and showed increased expression of the microglial activation marker MAC2 (Fig. 2c, d), indicative of a progressive neuroinflammatory response.

Adh5$^{-/-}$*Csb*$^{m/m}$ mice have a shortened lifespan compared with wild-type and *Csb*$^{m/m}$ mice. They live as long as *Adh5*$^{-/-}$ mice, but the cause of death between genotypes differs (Fig 2e, Extended Data Fig. 4a, b). The majority of *Adh5*$^{-/-}$ mice succumb to liver and lymphoid malignancies[16]; by contrast, *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice display a progressive decline in body condition and small nodular kidneys associated with biochemical markers of kidney dysfunction (Fig. 2f, g, Extended Data Fig. 4c). Histological analysis revealed that kidneys of young (3-month-old) *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice showed mild chronic inflammation, consistent with focal tubule loss progressing to mild chronic interstitial fibrosis with tubular atrophy and inflammatory infiltration by middle age (1-year-old). The changes in the terminal kidneys are consistent with overlapping features of ongoing kidney injury to the proximal tubules and end-stage renal failure (Fig. 2h, Extended Data Fig. 4d–f). The absence of glomerular changes before end-stage kidney failure is consistent with the lack of proteinuria (Extended Data Fig. 4g), and anaemia was associated with end-stage renal failure in *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice (Extended Data Fig. 4h). Detailed analysis of these mice did not reveal evidence of malignancies in their livers or secondary lymphoid organs. We did, however, observe mild age-related increase in ploidy and evidence of liver dysfunction (Extended Data Fig. 4i, j, k). Taken together, these data show that *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice ultimately succumb to kidney failure and have a reduced predisposition towards malignancies compared with *Adh5*$^{-/-}$ mice.

Levels of the formaldehyde DNA adduct, N^2-methyl-deoxyguanosine (N^2-Me-dG), a biomarker of exposure, were significantly elevated in brain and kidney in both *Adh5*$^{-/-}$ and *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice (Fig. 2i). Higher N^2-Me-dG levels were observed in male *Adh5*$^{-/-}$ mice compared with females, in line with the phenotypic sexual dimorphism. We next challenged mice with weekly intraperitoneal injections of 1.5 g kg^{-1} methanol, which is oxidized to formaldehyde (Extended Data Fig. 5a). This treatment led to a rapid decline in kidney function in *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice, resulting in kidney failure after 3–6 months of exposure (Extended Data Fig. 5b, c, d). *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice treated with methanol had small nodular kidneys and histology confirmed chronic end-stage kidney failure (Extended Data Fig. 5e, f). Brains of the treated mice showed a decrease in mass accompanied by the presence of MAC2+ microglial cells, and there was evidence of liver dysfunction (Extended Data Figs. 5g–i, 6a). Individuals with mutations in *XPA* do not develop Cockayne syndrome, although there is some phenotypic overlap[17]. Aged *Adh5*$^{-/-}$*Xpa*$^{-/-}$ mice displayed no features of Cockayne syndrome unless they were subjected to long-term methanol exposure, which resulted in chronic kidney failure (Extended Data Fig. 6b–d). We also tested whether the closely related aldehyde acetaldehyde could drive the Cockayne syndrome phenotype in *Aldh2*$^{-/-}$*Csb*$^{m/m}$ mice[18,19]—these mice

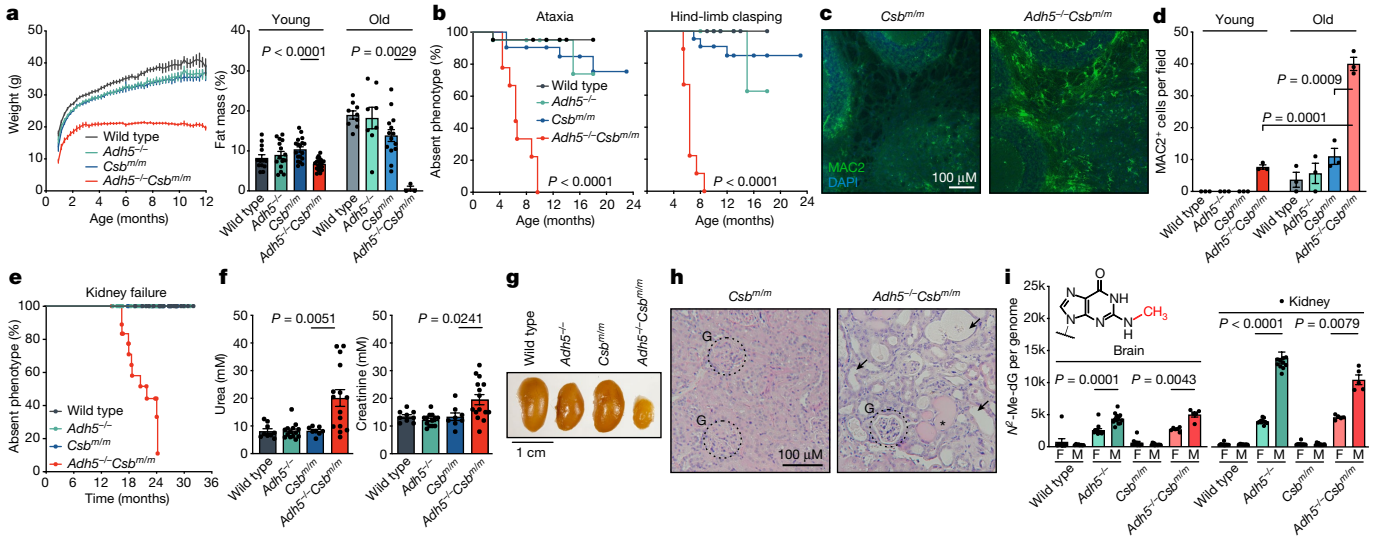


Fig. 2 | Endogenous formaldehyde accumulation reveals Cockayne syndrome. a, Left, growth of male mice based on weekly weights (data are mean ± s.e.m.; $n = 18$ (wild type), 15 ($Adh5^{-/-}$), 18 ($Csb^{m/m}$) and 20 ($Adh^{-/-}Csb^{m/m}$)) and fat mass measured at 11 weeks (young) and 53 weeks (old) (two-sided Mann–Whitney test; data are mean ± s.e.m.; $n = 13$ (young wild type), 15 (young $Adh5^{-/-}$), 18 (young $Csb^{m/m}$), 19 (young $Adh5^{-/-}Csb^{m/m}$), 9 (old wild type), 8 (old $Adh5^{-/-}$), 14 (old $Csb^{m/m}$) and 3 (old $Adh5^{-/-}Csb^{m/m}$)). **b**, Age of onset plots for ataxia and hind-limb clasping phenotypes (two-sided Mantel–Cox log-rank test; $n = 21$ (wild type), 19 ($Adh5^{-/-}$), 22 ($Csb^{m/m}$) and 9 ($Adh^{-/-}Csb^{m/m}$)). **c**, Representative immunofluorescence of MAC2 in the cerebellum of $Csb^{m/m}$ and $Adh5^{-/-}Csb^{m/m}$ mice. **d**, The number of MAC2+ cells per field in young (3-month-old) and old (18-month-old) mice (two-tailed Student's *t*-test; data are mean ± s.e.m.; $n = 3$ mice). **e**, Kidney failure-free survival plot ($n = 18$, 31, 26 and 19). **f**, Terminal serum urea and creatinine levels (two-sided Mann-Whitney test; data shown as mean and s.e.m.; $n = 9$ (wild type), 14 ($Adh5^{-/-}$), 8 ($Csb^{m/m}$) and 15 ($Adh^{-/-}Csb^{m/m}$)). **g**, Representative images of kidneys from mice at experimental end point. **h**, Representative haematoxylin and eosin-stained sections of kidneys of mice of indicated genotype. G indicates glomeruli arrows indicate atrophic tubules, and the asterisk shows evidence of tubular thyroidization. **i**, The frequency of N^2-Me-dG per genome in female (F) and male (M) mice. The inset shows the structure of the formaldehyde N^2-Me-dG adduct (two-sided Mann–Whitney test; data are mean ± s.e.m; $n = 9$ (wild-type brain), 8 ($Adh5^{-/-}$ brain), 12 ($Csb^{m/m}$ brain), 12 ($Adh^{-/-}Csb^{m/m}$ brain), 9 (wild-type kidney), 13 ($Adh5^{-/-}$ kidney), 6 ($Csb^{m/m}$ kidney) and 5 ($Adh^{-/-}Csb^{m/m}$ kidney)).

displayed no overt phenotype, reduced lifespan or signs of a Cockayne syndrome phenotype, even when continuously exposed to ethanol (Extended Data Fig. 6e–g). In summary, accumulation of formaldehyde from endogenous or exogenous sources is sufficient to precipitate the central features of the human Cockayne syndrome phenotype in mice.

Kidney cells susceptible to formaldehyde

Why the kidney fails in Cockayne syndrome is poorly understood. However, recent advances in single-cell RNA sequencing (scRNA-seq) have enabled identification of cell types susceptible to particular disease states[20]. We used scRNA-seq to profile nephron cellular responses to formaldehyde. Kidneys were dissociated to generate single-cell suspensions and sequenced using the 10x Genomics platform (Extended Data Fig. 7a). Uniform manifold approximation and projection (UMAP) of the data showed distinct nephron epithelial cell clusters, as well as several immune cell clusters, which were annotated on the basis of canonical marker expression (Fig. 3a, Extended Data Fig. 7b, c). When considering all kidney cell types, proximal tubule cells showed the greatest number of differentially expressed genes (DEGs) in $Adh5^{-/-}Csb^{m/m}$ kidneys compared with wild-type counterparts (Fig. 3b). Additionally, these cells partitioned to a specific region of the proximal tubule cluster (Extended Data Fig. 7d). To investigate this further, we interrogated the proximal tubule cells in isolation and found 8 sub-clusters (PT-0 to PT-7, Fig. 3c). Notably, the PT-4 sub-cluster was markedly expanded in $Adh5^{-/-}Csb^{m/m}$ kidneys (Fig. 3d,e, Extended Data Fig. 7e). These data thus identify proximal tubule cells as those with the most altered transcriptional profile in $Adh5^{-/-}Csb^{m/m}$ kidneys.

Damaged proximal tubule cells express GDF15

To address how the differences in gene expression in $Adh5^{-/-}Csb^{m/m}$ proximal tubule cells relate to their altered state and function, we ranked the top 100 marker genes for each proximal tubule sub-cluster (Supplementary Table 1). Among the marker genes for PT-4 we identified several indicative of kidney injury[21–23] (*Kim-1*, *Spp1*, *Cyr61* and *B2m*), with *Kim-1* being the top marker gene for this cluster and expression being restricted to those cells from $Adh5^{-/-}Csb^{m/m}$ proximal tubules (Fig. 4a, Extended Data Figs. 7f, g, 8a). Kidney injury molecule 1 (KIM-1) is a membrane protein that is induced in proximal tubule cells in response to chemotherapeutic agents such as cisplatin, and is an established clinical biomarker of proximal tubule damage[22,24]. This expression profile is consistent with the conclusion that the PT-4 sub-cluster represents damaged proximal tubule cells. To confirm these transcriptional changes, we carried out immunostaining of KIM-1, revealing expression in tubular cells only in $Adh5^{-/-}Csb^{m/m}$ mice.(Fig. 4b, Extended Data Fig. 7g).

STRING analysis of the top 100 marker genes of PT-4 identified five major nodes that enriched for distinct gene ontology (GO) terms (Extended Data Fig. 8b). One of the enriched pathways was alcohol metabolic processes, suggesting that these cells may be adapting to loss of ADH5. Additionally, there was an enrichment of genes that regulate apoptosis, a process triggered to promote clearance of damaged or dysfunctional cells. One mechanism by which proximal tubule cells clear apoptotic cells is through phagocytosis; of note, a recent study indicated that KIM-1-dependent phagocytosis can trigger major histocompatibility complex class II (MHCII) antigen presentation[24], and in line with this we observed increased expression of many genes involved in antigen presentation, particularly MHCII genes (Extended Data Fig. 8b). Tubular injury is also known to result in inflammation, including type I interferons (IFN), and type I IFN receptor-deficient mice are protected from tubular necrosis[25]. Notably, PT-4 cells showed a marked induction of interferon β response genes. Finally, there was an enrichment of protein activation cascade genes in PT-4 cells, which included not only markers of kidney injury (*Kim-1*, *Cyr61* and *Spp1*), but

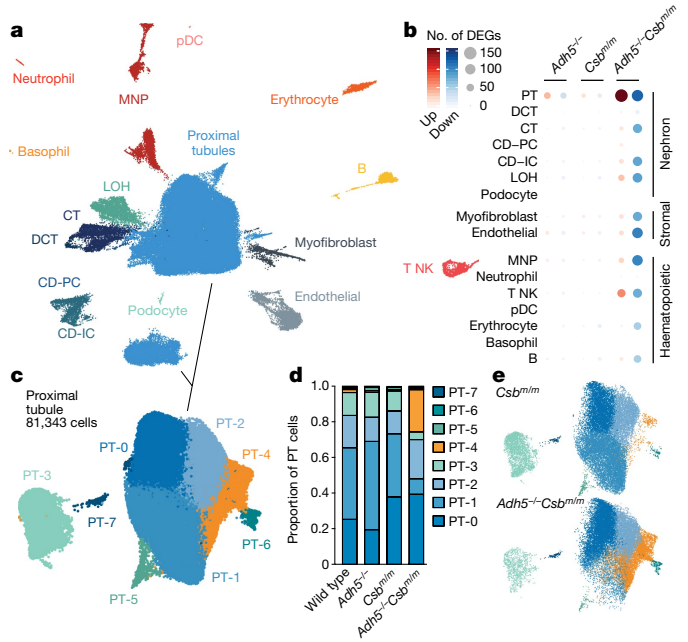


Fig. 3 | Single cell RNA-seq identifies regions in the nephron that are susceptible to transcriptional stress. a, UMAP plot of 106,531 mouse kidney cells. **b**, Number of DEGs that are up-regulated (red) or down-regulated (blue) relative to wild-type cells **c**, UMAP plot of 81,343 proximal tubule cells. **d**, The proportion of cells in each proximal tubule sub-cluster across all genotypes. **e**, UMAP plot of proximal tubule sub-clusters split by genotype. CT, connecting tubule; CD-IC, collecting duct—intercalated cell; CD-PC, collecting duct—principle cell; DCT, distal convoluted tubules; LOH, loop of Henle; MNP, mononuclear phagocyte; pDC, plasmacytoid dendritic cell; T NK, T and natural killer cells; B, B cells.

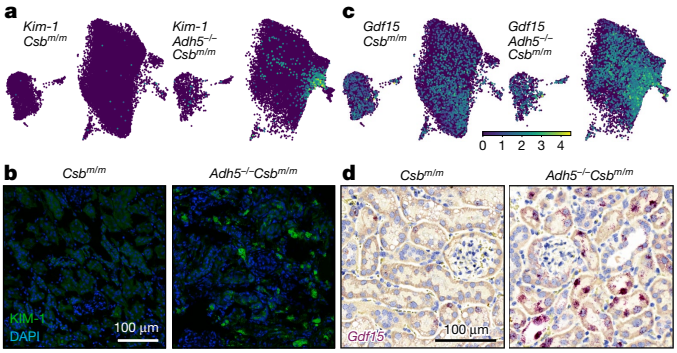


Fig. 4 | Transcriptional stress damages a subset of proximal tubule cells that express GDF15. a, Feature plots of *Kim-1* (also known as *Havcr1*) expression from proximal tubule sub-clusters. **b**, Representative immunofluorescence of KIM-1 in kidney sections. **c**, Feature plots of *Gdf15* expression in proximal tubule sub-clusters. **d**, Representative in situ hybridization of *Gdf15* mRNA (red) in kidneys from mice of the indicated genotype.

also fibrinogen genes, whose expression is increased in kidney injury and may have a role in remodelling damaged tissue[26].

When considering DEGs in cells within the PT-4 cluster, *Adh5*$^{-/-}$*Csb*$^{m/m}$ cells showed increased expression of a number of genes, including *p21* and *Gdf15*, which are known to be regulated by p53 (Extended Data Fig. 7h). Expression of both these genes was increased in the PT-4 cluster of *Adh5*$^{-/-}$*Csb*$^{m/m}$ cells, but they were expressed at low levels in control proximal tubule cells (Fig. 4c, Extended Data Figs. 7i, 8c). The p21 protein is a cell cycle inhibitor and DNA damage checkpoint protein[27,28]. *Gdf15* encodes a protein of the transforming growth factor-β (TGF-β) superfamily that is a well-known appetite suppressor[29,30]. We also identified additional p53 target genes that were induced in *Adh5*$^{-/-}$*Csb*$^{m/m}$ proximal tubule cells (Extended Data Fig. 8c, d), indicating that expression of *Gdf15* and *p21* is likely to be a consequence of a generalized p53 transcriptional response in these cells. Finally, we confirmed expression of *Gdf15* mRNA in the proximal tubule cells of *Adh5*$^{-/-}$*Csb*$^{m/m}$ by RNA in situ hybridization (Fig. 4d, Extended Data Fig. 7i). Taken together, kidney scRNA-seq indicates that *Adh5*$^{-/-}$*Csb*$^{m/m}$ proximal tubule cells are damaged by endogenous formaldehyde and respond by activating a p53 response which includes expression of the anorexic hormone GDF15.

GDF15 mediates DNA damage-induced cachexia

A notable feature of Cockayne syndrome is severe growth retardation and cachexia[5]. Patients with Cockayne syndrome have poor appetites and often require enteral feeding tubes[7,31]. GDF15 is an anorexic hormone, and its receptor GFRAL is restricted to a region in the brainstem important for feeding behaviour[32]. *Gdf15* mRNA levels were increased in the proximal tubule cells of kidneys of *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice; this corresponded with increased serum levels of GDF15 (Fig. 5a). Of note, we found a 4.5- and 1.9-fold increase in the serum GDF15 levels in male and female *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice, respectively. This correlates with our observation that *Adh5*$^{-/-}$*Csb*$^{m/m}$ males displayed more marked age-dependent loss of fat compared with females (Fig. 2a, Extended Data Fig. 3a). We next tested whether methanol treatment could induce GDF15 in females and precipitate weight loss. Indeed, methanol exposure led to a 6.7-fold increase in serum GDF15 in *Adh5*$^{-/-}$*Csb*$^{m/m}$ females, 13-fold higher than untreated wild-type controls (Fig. 5b). This was coupled with induction of GDF15 expression in the proximal tubule cells of methanol-exposed *Adh5*$^{-/-}$*Csb*$^{m/m}$ females (Fig. 5c) and weight loss (Fig. 5d).

While our data shows that proximal tubule cells express GDF15, we wanted to assess whether other organs also contributed to the serum levels of GDF15 observed in *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice. We observed no expression of GDF15 in other tissues known to express this hormone, such as the colon, small intestine, skeletal muscle and liver (Extended Data Fig. 9a). We did, however, observe some expression in kidney medullary cells in addition to proximal tubule cells (Extended Data Fig. 9b), which was also reflected in our scRNA-seq datasets (Extended Data Fig. 9c–f). Although the GDF15–GFRAL axis is a well-established anorexic endocrine circuit that can contribute to weight loss[29,32], *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice may be lean owing to increased energy expenditure. Calorimetry and open field tests revealed a mild increase in respiratory exchange rate but no increase in energy expenditure or activity levels (Extended Data Fig. 10a–c). Food intake, however, was reduced in *Adh5*$^{-/-}$*Csb*$^{m/m}$ males (Extended Data Fig. 10d). Taken together, these data suggest that the kidney is the key organ of GDF15 secretion in *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice, although we cannot exclude other sites that we did not survey.

GDF15 is also released after exposure to the chemotherapeutic agent cisplatin[32,33], but the molecular mechanism of its secretion remains unclear. Similar to formaldehyde, cisplatin induces DNA damage that is repaired by TC-NER[34]. We exposed wild-type and *Csb*$^{m/m}$ mice to cisplatin and monitored weight and serum GDF15 levels (Extended Data Fig. 11a). Wild-type mice exposed to a low dose of cisplatin displayed no weight loss or GDF15 induction. By contrast, *Csb*$^{m/m}$ mice rapidly lost weight and showed a fivefold induction of serum GDF15. Next, we tested whether cisplatin-induced GDF15 release was dependent on p53. A higher dose of cisplatin supressed weight gain in wild-type mice and also induced serum GDF15 by 12-fold —by contrast, *p53*$^{-/-}$ mice continued to gain weight and correspondingly showed no induction of serum GDF15 (Extended Data Fig. 11b) or GDF15 mRNA in proximal tubule cells (Extended Data Fig. 11c).

To establish that GDF15 contributes to progressive weight loss in *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice, we tested the consequences of blocking this

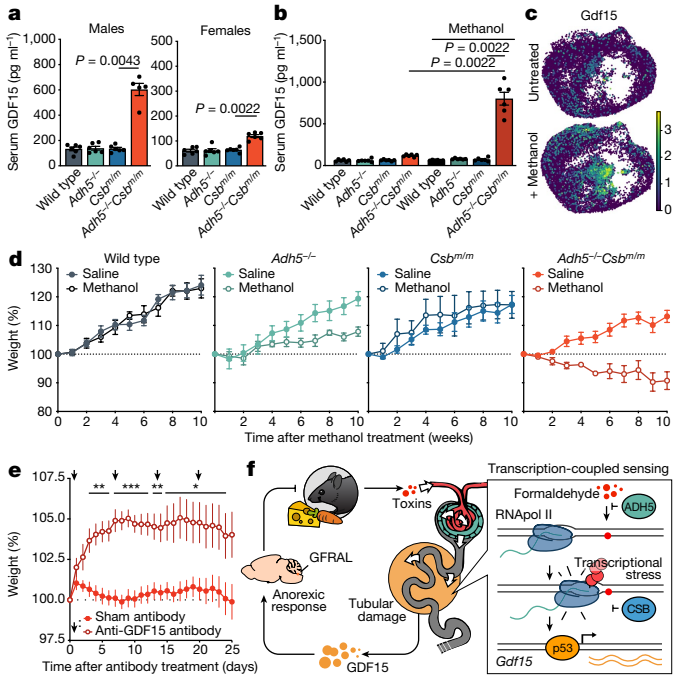


Fig. 5 | GDF15 mediates DNA damage induced cachexia. a, Serum GDF15 levels in male (left) and female (right) mice (two-sided Mann–Whitney test; data are mean ± s.e.m.; $n = 6$ (wild type), 6 ($Adh5^{-/-}$), 6 ($Csb^{m/m}$) and 5 ($Adh^{-/-}Csb^{m/m}$) for males and $n = 6$ (all genotypes) for females). **b**, Serum GDF15 levels in female mice following 5 weeks of treatment with methanol or saline (two-sided Mann–Whitney test; data are mean ± s.e.m.; $n = 6$ per genotype). **c**, Feature plots of *Gdf15* expression in proximal tubule cells of $Adh5^{-/-}Csb^{m/m}$ female mice exposed to methanol or saline (untreated). **d**, Body weight of female mice exposed to methanol or saline (data are mean ± s.e.m.; $n = 4$ per genotype). **e**, Seven-day rolling average percentage weight change of $Adh5^{-/-}Csb^{m/m}$ mice after treatment with a neutralizing GDF15 monoclonal antibody or isotype control (sham) (two-sided Mann–Whitney test (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$); $n = 7$ per treatment); data are mean ± s.e.m.). Arrows denote days of treatment. **i**, Left, model outlining how formaldehyde damages proximal tubules and triggers release of the anorexic hormone GDF15, resulting in food aversion. Left, model of the mechanism by which ADH5 and CSB protect cells from formaldehyde damage, which is sensed by the stalled transcription machinery and can trigger p53-dependent activation of *Gdf15*.

response. Male $Adh5^{-/-}Csb^{m/m}$ mice received a previously validated neutralising mouse monoclonal antibody against GDF15 (Mab-1) or an isotype control[33] weekly for one month. $Adh5^{-/-}Csb^{m/m}$ mice treated with Mab-1 showed a marked and sustained increase in body weight that equated to an average of 5% of the original weight (Fig. 5e). By contrast, the body weights of the cohort treated with control antibody remained unchanged. In summary, $Adh5^{-/-}Csb^{m/m}$ mice show a strong correlation between growth failure, methanol-induced weight loss and serum levels of GDF15. Thus, transcription-blocking DNA damage activates p53-dependent GDF15 secretion, and neutralising this response attenuates cachexia.

Discussion

In this study, we found that endogenously-produced formaldehyde is sufficient to reveal Cockayne syndrome in mice. Using this model, we identified a cellular target for formaldehyde-induced transcriptional stress within the nephron and showed that these cells trigger an endocrine weight loss response (Fig. 5f). Formaldehyde also precipitates Fanconi anaemia in mice[12]. In that model, highly proliferative cell types (for example, haematopoietic tissues) use replication-coupled Fanconi anaemia DNA repair to resolve formaldehyde damage[35]. By contrast, cells that are either post-mitotic (for example, neurons) or divide infrequently (for example, nephrons and hepatocytes) may be more dependent on a transcription-coupled response. Formaldehyde exposure leads to the stalling of Pol II, activation of TC-NER and the degradation of Pol II, perhaps by causing DNA intra-strand crosslinks. UV and cisplatin also cause such crosslinks, which are repaired by TC-NER[1,34,36]. Both XPA and CSB are required for cellular resistance to formaldehyde, suggesting that the canonical TC-NER machinery excises DNA damage caused by this aldehyde. However, in vivo, the phenotype of $Adh5^{-/-}Xpa^{-/-}$ mice is milder than that of $Adh5^{-/-}Csb^{m/m}$ mice, analogous to the phenotype of human patients with mutations in *XPA* and *CSB*[5,7,17]. The discordance in severity of the mouse phenotypes could be because CSB protects Pol II from degradation in response to formaldehyde. This function has been proposed to explain the severity of the phenotype in human patients with Cockayne syndrome[4]. Alternatively, CSB may facilitate lesion bypass, similar to the yeast homologue of CSB (Rad26), using Pol II forward-translocating activity[2].

Cisplatin is nephrotoxic and stimulates GDF15 secretion in a p53-dependent manner, stimulating anorectic weight loss. Blocking the GDF15–GFRAL axis may alleviate the cachexia associated with chemotherapy[33], and such an approach may extend to alleviating cachexia in Cockayne syndrome. The GDF15 response in proximal tubule cells may have evolved in mammals to ensure aversion to genotoxic food sources. In Cockayne syndrome, this response is constitutively active because of the absence of relief from transcriptional stress. It is plausible that this response is appropriate to avoid dietary sources of endogenous formaldehyde. Support for nutritional genotoxicity comes from the attenuation of neurodegeneration and kidney abnormalities in caloric-restricted mice that lack the TC-NER nuclease, XPF–ERCC1[37]. Finally, genome-wide association studies conducted on patients with anorexia nervosa identified four susceptibility loci, one of which is the locus encoding the DNA repair enzyme MGMT[38]. The induction of GDF15 by DNA damage may therefore have broader implications for other human anorectic syndromes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04133-7.

1. Brueckner, F., Hennecke, U., Carell, T. & Cramer, P. CPD damage recognition by transcribing RNA polymerase II. *Science* **315**, 859–862 (2007).
2. Xu, J. et al. Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature* **551**, 653–657 (2017).
3. Gregersen, L. H. & Svejstrup, J. Q. The cellular response to transcription-blocking DNA damage. *Trends Biochem. Sci.* **43**, 327–341 (2018).
4. Tufegdžić Vidaković, A. et al. Regulation of the RNAPII pool is integral to the DNA damage response. *Cell* **180**, 1245–1261.e21 (2020).
5. Laugel, V. Cockayne syndrome: the expanding clinical and mutational spectrum. *Mech. Ageing Dev.* **134**, 161–170 (2013).
6. Stern-Delfils, A. et al. Renal disease in Cockayne syndrome. *Eur. J. Med. Genet.* **63**, 103612 (2020).
7. Kubota, M. et al. Nationwide survey of Cockayne syndrome in Japan: incidence, clinical course and prognosis. *Pediatr. Int.* **57**, 339–347 (2015).
8. Lans, H., Hoeijmakers, J. H. J., Vermeulen, W. & Marteijn, J. A. The DNA damage response to transcription stress. *Nat. Rev. Mol. Cell Biol.* **20**, (2019).
9. Wilson, M. D., Harreman, M. & Svejstrup, J. Q. Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. *Biochim. Biophys. Acta* **1829**, 151–157 (2013).
10. van der Horst, G. T. et al. Defective transcription-coupled repair in Cockayne syndrome B mice is associated with skin cancer predisposition. *Cell* **89**, 425–435 (1997).
11. van der Horst, G. T. J. et al. UVB radiation-induced cancer predisposition in Cockayne syndrome group A (*Csa*) mutant mice. *DNA Repair* **1**, 143–157 (2002).
12. Pontel, L. B. et al. Endogenous formaldehyde is a hematopoietic stem cell genotoxin and metabolic carcinogen. *Mol. Cell* **60**, 177–188 (2015).
13. Cheng, G. et al. Reactions of formaldehyde plus acetaldehyde with deoxyguanosine and DNA: formation of cyclic deoxyguanosine adducts and formaldehyde cross-links. *Chem. Res. Toxicol.* **16**, 145–152 (2003).

14. Marteijn, J. A., Lans, H., Vermeulen, W. & Hoeijmakers, J. H. J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **15**, 465–481 (2014).
15. Ranes, M. et al. A ubiquitylation site in Cockayne syndrome B required for repair of oxidative DNA damage, but not for transcription-coupled nucleotide excision repair. *Nucleic Acids Res.* **44**, 5246–5255 (2016).
16. Wei, W. et al. *S*-nitrosylation from GSNOR deficiency impairs DNA repair and promotes hepatocarcinogenesis. *Sci. Transl. Med.* **2**, 19ra13 (2010).
17. Niedernhofer, L. J., Bohr, V. A., Sander, M. & Kraemer, K. H. Xeroderma pigmentosum and other diseases of human premature aging and DNA repair: Molecules to patients. *Mech. Ageing Dev.* **132**, 340–347 (2011).
18. Langevin, F., Crossan, G. P., Rosado, I. V., Arends, M. J. & Patel, K. J. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature* **475**, 53–59 (2011).
19. Garaycoechea, J. I. et al. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature* **489**, 571–575 (2012).
20. Young, M. D. et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).
21. Feng, D., Ngov, C., Henley, N., Boufaied, N. & Gerarduzzi, C. Characterization of matricellular protein expression signatures in mechanistically diverse mouse models of kidney injury. *Sci. Rep.* **9**, 16736 (2019).
22. Vaidya, V. S., Ramirez, V., Ichimura, T., Bobadilla, N. A. & Bonventre, J. V. Urinary kidney injury molecule-1: a sensitive quantitative biomarker for early detection of kidney tubular injury. *Am. J. Physiol.* **290**, 517–529 (2006).
23. Dieterle, F. et al. Urinary clusterin, cystatin C, B2-microglobulin and total protein as markers to detect drug-induced kidney injury. *Nat. Biotechnol.* **28**, 463–469 (2010).
24. Brooks, C. R. et al. KIM-1/TIM-1-mediated phagocytosis links ATG5/ULK1-dependent clearance of apoptotic cells to antigen presentation. *EMBO J.* **34**, 2441–2464 (2015).
25. Freitas, M. C. S. et al. Type I interferon pathway mediates renal ischemia/reperfusion injury. *Transplantation* **92**, 131–138 (2011).
26. Liu, J. et al. Molecular characterization of the transition from acute to chronic kidney injury following ischemia/reperfusion. *JCI Insight* **2**, (2017).
27. Espinosa, J. M., Verdun, R. E. & Emerson, B. M. p53 functions through stress- and promoter-specific recruitment of transcription initiation components before and after DNA damage. *Mol. Cell* **12**, 1015–1027 (2003).
28. Campisi, J. & D'Adda Di Fagagna, F. Cellular senescence: when bad things happen to good cells. *Nat. Rev. Mol. Cell Biol.* **8**, 729–740 (2007).
29. Mullican, S. E. et al. GFRAL is the receptor for GDF15 and the ligand promotes weight loss in mice and nonhuman primates. *Nat. Med.* **23**, 1150–1157 (2017).
30. Tufegdzic Vidakovic, A. et al. Analysis of RNA polymerase II ubiquitylation and proteasomal degradation. *Methods* **159–160**, 146–156 (2019).
31. Wilson, B. T. et al. The Cockayne Syndrome Natural History (CoSyNH) study: clinical findings in 102 individuals and recommendations for care. *Genet. Med.* **18**, 483–493 (2016).
32. Hsu, J. Y. et al. Non-homeostatic body weight regulation through a brainstem-restricted receptor for GDF15. *Nature* **550**, 255–259 (2017).
33. Breen, D. M. et al. GDF-15 neutralization alleviates platinum-based chemotherapy-induced emesis, anorexia, and weight loss in mice and nonhuman primates. *Cell Metab.* **32**, 938–950.e6 (2020).
34. Enoiu, M., Jiricny, J. & Schärer, O. D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res.* **40**, 8953–8964 (2012).
35. Knipscheer, P. et al. The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* **326**, 1698–1701 (2009).
36. Mulderigg, L. & Garaycoechea, J. I. XPF-ERCC1 protects liver, kidney and blood homeostasis outside the canonical excision repair pathways. *PLoS Genet.* **16**, e1008555 (2020).
37. Vermeij, W. P. et al. Restricted diet delays accelerated ageing and genomic stress in DNA-repair-deficient mice. *Nature* **537**, 427–431 (2016).
38. Watson, H. J. et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Methods

Cell culture

tMEFS were cultured in standard Dulbecco's modified Eagle's medium (DMEM) (GIBCO) supplemented with 10% fetal calf serum (GIBCO) and penicillin-streptomycin. Flp-In T-REx HEK 293T cell lines (Thermo Fisher Scientific) were cultured in standard DMEM supplemented with 10% fetal calf serum and penicillin-streptomycin. CSB KO HEK 293 cells were a gift from J. Svejstrup (Francis Crick Institute).

Generation of tMEFs

Adh5$^{+/-}$*Xpc*$^{+/-}$, *Adh5*$^{+/}$*Xpa*$^{+/-}$ and *Adh5*$^{+/-}$*Csb*$^{+/m}$ mice were timed mated and humanely killed at embryonic day 13.5. Primary MEF cultures were obtained and transformed using pBABE-SV40-Puro virus as described[39].

For the complementation of tMEF lines, wild-type and K991R *Csb* cDNA constructs were amplified with a N-terminal 2×Flag tag from mouse cDNA. They were then sequenced and cloned into pExpress and subsequently into pLoxBsr. The plasmids were transfected into tMEF using Lipofectamine LTX (Invitrogen) following the manufacturer's instructions. Positive clones were selected using 10 μg ml^{-1} blasticidin and screened for expression by anti-Flag western blot. Selection was maintained on positive clones.

CRISPR–Cas9-mediated gene disruption of *ADH5* in HEK 293T cells

Wild-type and CSB KO HEK 293 cells were targeted using the top oligonucleotide sequence CACCGATCACTGGAAAACAACCCTC and bottom oligonucleotide sequence AAACGAGGGTTGTTTTCCAGTGATC cloned into pX458. Cells were transfected with Turbofectin 8.0 (Origene). Two days after transfection, GFP-positive cells were single-cell sorted into 96-well plates containing medium supplemented with 20% fetal calf serum, using a MoFlo cell sorter (Beckman–Coulter). After 14 days of incubation at 37 °C, individual clones were analysed for expression of ADH5 protein by western blotting.

Cytotoxicity assays

One-thousand cells (tMEFS or HEK 293) were plated in a 96-well plate in triplicate for each condition. The cells were exposed to the indicated doses of formaldehyde (Thermo Fisher Scientific) or UV 24 h after plating. Cell viability was measured using MTS reagent (Promega), 4 days after plating. Absorbance at 492 nm was measured with a Pherastar spectrophotometer.

Pol II RPB1 ubiquitylation

Detection of ubiquitylated RPB1 was performed as previously described in detail[30]. In brief, Dsk2 beads were prepared by transfecting One Shot BL21 competent cells with pGEX3-Dsk2, inducing them, collecting protein and adding the protein to glutathione Sepharose beads. Human cell lysates were also prepared as previously described[30], 30 min after exposure to UV or the indicated concentration of formaldehyde. Cell lysates and Dsk2 beads were then incubated together overnight before analysis by western blot.

Western blotting

For ADH5 and Flag-tag blots, cells were lysed for 30 min on ice in RIPA buffer (Thermo Fisher Scientific), including protease and phosphatase inhibitor cocktail (Roche) and phosphatase inhibitor cocktail (Roche). Protein samples were run on a 4–12% Bis-Tris gel (Thermo Fisher Scientific) and blotted onto a 0.45 μm nitrocellulose membrane. For Pol II RPB1 ubiquitylation and degradation, blotting was performed as previously described[30].

Antibodies

For screening HEK 293 cells, we used an in-house rabbit anti-ADH5 antibody that has been described previously[12]. For screening complemented tMEF clones, we used anti-Flag (Sigma, clone M2) at 1:1,000. For blots showing RPB1 degradation, we used anti-RPB1 (Cell Signalling, D8L4Y) at 1:1000 and for ubiquitylation of RPB1, we used RPB1 (abcam, 4H8) at 1:10,000. For both RPB1 blots, we used mouse anti-vinculin (Sigma, Cat V9131) as a loading control at 1/10,000.

Mice

Animal experiments undertaken in this study were with approval of the MRC Laboratory of Molecular Biology animal welfare and ethical review body and the UK Home Office under the Animal (Scientific Procedures) Act 1986 license PFC07716E.

All mice were maintained under specific pathogen-free conditions in individually ventilated cages (Techniplast GM500, Techniplast) on Ligno-cel FS14 spruce bedding (IPS, LTD) with environmental enrichment (fun tunnel, chew stick and Enviro-Dri nesting material (LBS)) at 19–23 °C with light from 07:00 to 19:00 and fed Dietex CRM pellets (Special Diet Services) ad libitum.

Xpa^{tm1Hvs} (MGI 1857939, C57BL/6), *Ercc6*^{tm1Guh} (*Csb*^{m}, MGI 1932102, C57BL/6) and *Xpc*^{tm1Ecf} (MGI 1859840, C57BL/6) mice were described previously[10,40,41] and were a gift from G. T. van der Horst (Erasmus University), E. Friedberg (UT Southwestern) and J. Hoeijmakers (Erasmus University). *Adh5*^{tm1Stam} (also known as *Gsnor*$^{-/-}$) (MGI 3033711, C57BL/6) mice were previously described[42] and obtained from L. Liu at UCSF. The *Aldh2* mice (*Aldh2*$^{tm1a(EUCOMM)Wtsi}$; MGI ID: 4431566, C57BL/6) were reported previously[18] and the *Trp53* allele that has been reported previously[43] was used in a C57BL/6 × 129S4S6/Sv hybrid background.

To generate *Adh5*$^{-/-}$*Xpa*$^{-/-}$, *Adh5*$^{-/-}$*Csb*$^{m/m}$ and *Adh5*$^{-/-}$*Xpc*$^{-/-}$ mice, *Adh5*$^{+/-}$ mice were crossed with *Xpa*$^{+/-}$, *Csb*$^{+/m}$ and *Xpc*$^{+/-}$, respectively. From the resulting progeny we inter-crossed *Adh5*$^{+/-}$NER$^{+/-}$ mice to generate all possible genotypes. To further bias the breeding for double homozygous mice we inter-crossed mice that were *Adh5*$^{-/-}$ NER$^{+/-}$ or *Adh5*$^{+/-}$ NER$^{-/-}$. Similarly, to generate *Aldh2*$^{-/-}$*Csb*$^{m/m}$ mice we crossed *Csb*$^{+/m}$ mice with *Aldh2*$^{+/-}$ and then from resulting progeny we inter-crossed *Aldh2*$^{+/-}$*Csb*$^{+/m}$ mice to generate all possible genotypes. To further bias the breeding for double homozygous mice we inter-crossed mice that were *Aldh2*$^{+/-}$*Csb*$^{m/m}$ or *Aldh2*$^{-/-}$*Csb*$^{+/m}$.

Cerebellar ataxia scoring was performed as previously described[44] at monthly intervals once the mice were 3 months old.

For methanol treatment, mice received 1.5 g kg^{-1} methanol via intraperitoneal injection once a week until the conclusion of the experiment. Body weights were recorded daily and 50-μl blood samples were taken monthly via tail bleed for serum biochemistry assays.

For ethanol exposure, the drinking supply was replaced with a solution of 10:20:80 blackcurrant Ribena:ethanol:water. Body weights were recorded daily and 50-μl blood samples were taken monthly via tail bleed for serum biochemistry assays.

For low-dose cisplatin exposure, 50-μl blood samples were taken before 0.5 mg kg^{-1} weekly cisplatin intraperitoneal injections. Mice were weighed daily throughout before terminal blood sampling after 4 weeks. For high-dose exposure, 4 mg kg^{-1} cisplatin was used for 6 weeks.

For GDF15 monoclonal antibody treatment experiments, *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice were treated with 10 mg kg^{-1} monoclonal antibody or IgG control by subcutanenous injection once a week for 4 weeks. The GDF15 monoclonal antibody and IGg control were provided by Pfizer.

Mouse phenotyping

The in vivo phenotyping tests below were carried out at MRC Harwell and more detailed descriptions of these protocols are available on the IMPRESS website (https://www.mousephenotype.org/impress). All procedures conducted were done so in accordance with the Animals (Scientific Procedures) Act 1986 Amendment Regulations 2012 (SI 4 2012/3039). Mice were housed in Tecniplast IVC cages (1284L and 1285L) with Aspen bedding (Datesand) which was autoclaved (134 °C for 3 min) in the cage base and cage bases were changed weekly Food and water were provided ad libitum. RM3 (E) Diet (Special Diet Services) was irradiated

to 2.5 Mrad. Mains water was supplied through a reverse-osmosis filter and then chlorinated to 9–13 ppm. Water bottles were changed weekly with autoclaved bottle and cap. The lighting regime was 12 h light: 12 h dark with 30 min dusk to dawn and dawn to dusk periods.

Body weight

Mice were weighed using a dynamic weighing balance (Ohaus) and were weighed in the morning to limit the effect of the circadian rhythm on body weight.

X-ray

X-ray images of the mice were collected whilst the mice were anaesthetized (isoflurane). The X-rays were captured using the Faxitron Ultrafocus X-Ray machine.

Grip strength

Grip strength was measured using a grip-strength meter (Bioseb), recording the maximum force generated by a mouse using all four limbs. Grip strength measures were carried out in triplicate for each mouse.

Calorimetry

The metabolic rate of the mice was assessed using indirect calorimetry. Mice were individually housed overnight for a period of 21 h in Phenomaster cages (TSE Systems) with standard bedding and igloos. The air content in each cage was sampled for 1 min, in turn, and a reference cage with no mice in it was also sampled for comparison. The oxygen consumption and carbon dioxide production for each mouse was then determined as the difference between the reference amount of gas and the level in the mouse cage.

Open field

Mice were transported to the testing room 30 min before the test. Lighting conditions were set between 150 and 200 lux. Animals were place in a chamber for 20 min and activity levels were determined using a beam-break system.

EchoMRI

Body composition of the mice was assessed using an EchoMRI whole-body composition analyser (Echo Medical System). The analysis output quantified fat mass, lean mass and water content of the mice.

Serum biochemistry

Serum was collected from 200 µl of whole blood into Microvette 200 conical tubes (MCV200-SER) after centrifugation. Levels of urea, creatinine, aspartate aminotransferase, albumin and alkaline phosphatase of serum samples were measured using a Siemens Dimension RxL analyser by Cambridge Biochemical Assay Laboratory. GDF15 was measured using a commercial ELISA (R&D Systems) according to the manufacturer's instructions with two modifications. We used an extended standard curve extending from 2,000 pg ml^{-1} to 7.8 pg ml^{-1}, and added 10 µl sample to 50 µl RD1W at the binding step (instead of 50 µl). Additional samples for GDF15 were measured by the Cambridge Biochemical Assay Laboratory, University of Cambridge using a Mouse GDF15 DuoSet ELISA (R&D Systems) which had been modified to run as an electrochemiluminescence assay on the Meso Scale Discovery assay platform.

Urine analysis

Multistix SG10 from Siemens was used for semi-quantitative readout of protein in the urine by addition of 5 µl urine to the strip.

Blood counts

Total blood was collected in K_3EDTA MiniCollect tubes (Greiner Bio-one) and analysed on a scil VetABC Plus+ blood counter (Horiba).

Histology

Histological analysis was performed on tissues that had been fixed in neutral buffered formalin for 24 h. The samples were embedded in paraffin and 4-µm sections were cut before staining with haematoxylin and eosin.

Immunofluorescence

Mice were anaesthetized with pentobarbital and perfused transcardially with 4% paraformaldehyde. The brain was carefully dissected out and post-fixed overnight in 4% paraformaldehyde. Brain tissue was then sectioned at 40 µm with a vibrating microtome and sections were processed, free floating. Sections were washed in PBST (PBS 0.1% Triton X) for 10 min before blocking for 1 h in 5% normal goat serum. Next, sections were incubated with MAC2 (Cedarlane, clone M3/38) antibody at 1:2,000 overnight at 4 °C before being washed once in PBST for 10 min and stained with Alexaflour 488 anti-rat secondary (Invitrogen) for 1 h at 1/1000 in the dark at room temperature. Sections were then transferred to slides and mounted using one drop of Vectashield containing DAPI (Vectorlabs) before cover slipping and adding nail varnish. Images were taken on a Zeiss LSM 780 confocal microscope (Zeiss). For kidney KIM-1 immunofluorescence, kidneys were snap frozen and 10 µm sections were cut using a freezing microtome and placed on slides, which were then stored at −80 °C. For staining, slides were air dried at room temperature for 5 min then fixed in 2% PFA for 5–10 min. For staining, slides were washed in PBS for 10 min before blocking in 10% fetal calf serum (GIBCO) at room temperature for 30 min in a humid chamber. KIM-1 primary antibody (R&D Systems, AF1817) was then added at 1:500 overnight at 4 °C in a humid chamber. The next day, slides were washed in PBS for 10 min before secondary anti-goat Alexa-fluor 488 (Invitrogen) was added at 1:1,000 for 30 min. Slides were then washed with PBS for 10 min before being mounted using one drop of Vectashield containing DAPI (Vectorlabs) before cover slipping and sealing with nail varnish. Images were taken on a Zeiss LSM 780 confocal microscope (Zeiss).

RNA in situ hybridization

Tissues for in situ hybridization were dissected and placed into 10% formalin/PBS for 24 h at room temperature, transferred to 70% ethanol and processed into paraffin. Five-micrometre sections were cut and mounted onto Superfrost Plus (Thermo Fisher Scientific). Detection of mouse *Gdf15* was performed on formalin-fixed paraffin-embedded sections using Advanced Cell Diagnostics (ACD) RNAscope 2.5 LS Reagent Kit-RED (no. 322150) and RNAscope LS 2.5 Probe Mm-Gdf15-O1 (no. 442948) (ACD). In brief, sections were baked for 1 h at 60 °C before loading onto a Bond RX instrument (Leica Biosystems). Slides were deparaffinized and rehydrated on board before pre-treatments using Epitope Retrieval Solution 2 (no. AR9640, Leica Biosystems) at 95 °C for 15 min, and ACD Enzyme from the LS Reagent kit at 40 °C for 15 min. Probe hybridization and signal amplification was performed according to the manufacturer's instructions. Fast red detection of mouse *Gdf15* was performed on the Bond RX using the Bond Polymer Refine Red Detection Kit (Leica Biosystems, no. DS9390) according to the ACD protocol. Slides were then counterstained with haematoxylin, removed from the Bond RX and were heated at 60 °C for 1 h, dipped in xylene and mounted using EcoMount Mounting Medium (Biocare Medical, no. EM897L). Slides were imaged on an automated slide-scanning microscope (Axioscan Z1 and Hamamatsu orca flash 4.0 V3 camera) using a 20× objective with a numerical aperture of 0.8.

Nuclei isolation and DNA content analysis

Livers were dissected and passed through a 40-µm filter. Cells were washed twice in LA buffer (250 mM sucrose, 5 mM $MgCl_2$ and 10 mM Tris-HCl, pH 7.4). After washing, the cell pellet was resuspended in 1 ml of buffer LB (2 M sucrose, 1 mM $MgCl_2$ and 10 mM Tris-HCl, pH 7.4) and centrifuged at 16,000*g* for 30 min. The white nuclei-containing pellet

was resuspended in LA buffer and kept on ice for analysis. For DNA content analysis, nuclei were fixed drop-wise in cold 96% ethanol. Nuclei were pelleted and re-suspended in 400 µl PBS. Propidium iodide solution (Sigma) was added at a final concentration of 40 µg ml^{-1} together with Ribonuclease A (Sigma) at a final concentration of 100 µg ml^{-1}. The samples were incubated on ice for 1 h and then analysed on LSRII flow cytometer (BD Pharmingen). The data were analysed with FlowJo 10.0.6 (Tree Star).

Synthesis of nucleoside standards

Chemistry and the standards used were as previously reported[45].

Tissue sample preparation for determination of N^2-Me-dG in DNA

Tissue was lysed and DNA was extracted using Puregene solutions, 4 µl of proteinase K and a 7 mm stainless steel metal ball. Samples where homogenized in a tissue lyser (Qiagen/Retsch) for 4 min at 30 Hz, then incubated at 37 °C for 30 min, then 4 µl of RNase A was added, vortexed and incubated at 37 °C for 1 h. Supernatant transferred (1.7-ml tube) and 266 µl protein precipitation solution was added and vortexed, cooled on ice for 5 min, the centrifuged at 21,300*g* for 3 min. Supernatant was transferred, and DNA precipitated with isopropanol and 70% ethanol. The pellet left to air dry for 5 min before being reduced in 500 µl of 50 mM $NaCNBD_3$, 200 mM NaOAc (pH 5.2) for 48 h at room temperature. DNA was precipitated with isopropanol and 70% ethanol and the pellet left to air dry and dissolved in 100 µl of ultra-pure water (Romil). Five thousand nanograms of DNA was digested with an enzyme cocktail in a total volume of 100 µl; also added were the internal standards ^{15}N-N^2-Me-dG and ^{15}N-dA. Digestion was allowed to proceed for at least 16 h, then samples were filtered with a 2,000 MWCO Vivacon 500 filter for 40 min at 16,000*g*. Samples were then transferred to a mass spectrometry vial and analysed.

Liquid chromatography–tandem mass spectrometry determination of N^2-Me-dG in DNA digests

Samples were analysed on a TSQ Altis Triple Quadrupole mass spectrometer (SRM mode) interfaced to an UltiMate 3000 uHPLC. Ten microlitres of sample (500 ng of digested DNA on the column) was injected per run using the full-loop inject mode. Buffers used were, A (H_2O (0.1% acetic acid)) and B (MeCN (0.1% acetic acid)).

Kidney dissociation for scRNA-seq

Kidneys were placed into a small tube containing 1 ml Digest mix (RPMI, 25 µg ml^{-1} Liberase TM, 50 µg ml^{-1} DNase) and minced to less than 2 mm^3. The volume was increased to 5 ml with digest mix and samples were incubated at 37 °C for 30 min in a shaking incubator. Samples were passed through a 100-µm cell strainer using a 1-ml syringe plunger and washed by centrifugation with PBS. Samples were resuspended in PBS, cell number determined and loaded onto the 10x Genomics platform.

scRNA-seq data analysis

Following sequencing, BCL files were demutiplexed to Fastq files using CASAVA. Subsequently splitting to single cells and mapping of genes was carried out using Cellranger.

The single-cell data (10x cellranger output) was corrected for ambient RNA expression using SoupX[46] (v1.2.1) . After SoupX, doublet detection was performed using scrublet[47] (v0.2.1) with adaptations outlined in Popescu et al.[48]. In brief, after processing in scrublet, the data were iteratively sub-clustered using the standard Seurat-inspired scanpy (v.1.4.5.post2) workflow[49,50] and a median scrublet score for each subcluster was computed. Median absolute deviation was computed from the cluster scrublet scores and a one tailed *t*-test was performed with Benjamini–Hochberg correction[51] applied where cells with significantly outlying cluster scrublet scores (Benjamini–Hochberg *P* value <0.1) were flagged as potential doublets. The data were then processed using scanpy with standard quality control steps; cells were filtered if the number of genes was greater than 2,500 or less than 200. Similar to Park et al.[52], the per cent mitochondrial content cut-off was set at <50%. Genes were retained if expressed by at least three cells. Gene counts for each cell were normalized to contain a total count equal to 10,000 per cell. Highly variable genes were selected based on the following parameters: minimum and maximum mean expression are ≥0.0125 and ≤3 respectively; minimum dispersion of genes = 0.5. The effects of total counts per cell, percentage of mitochondrial content and cell cycle states were regressed out prior to scaling of variable genes. Neighbourhood graph graphs were constructed with 20 neighbours using 80 principal components (all cells) or 30 principal components (proximal tubules only). Clustering was performed using the Leiden algorithm[53] with resolution set at 0.5. UMAP[54] (v3.10.0) was used for dimensional reduction and visualization; Partition-based graph abstraction initialization was used to compute the UMAP embedding for all cells, and default (spectral) was used for proximal tubules; the minimum distance was set at 0.3; all other parameters were used as per default settings in scanpy.

Differential gene testing was performed using the Wilcoxon test rank sum implemented in scanpy's rank_genes_groups module. Benjamini–Hochberg post hoc correction was applied to the *P* values calculated after DEG analysis. Genes that attained corrected *P* value <0.05 were considered statistically significant.

Statistical analysis

Sample number (*n*) indicates the number of independent biological samples in each experiment. Sample numbers and experimental repeats are indicated in figure legends or Methods. Unless otherwise stated in the figure legend, data are shown as the mean ± s.e.m. and the two-sided nonparametric Mann–Whitney test was used to assess statistical significance. Analysis was performed using GraphPad Prism.

Statistics and reproducibility

For Fig. 1e–g, blots were repeated at least twice. For Fig. 2c, the number of mice used is the same as in Fig. 2d. Figure 2h shows representative histology from the mice plotted in Fig. 2e and Fig. 4b is representative *n* = 3.

Reporting summary

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

Data availability

All scRNA-seq data have been deposited in the Gene Expression Omnibus under accession GSE175792.

39. Crossan, G. P. et al. Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. *Nat. Genet.* **43**, 147–152 (2011).
40. de Vries, A. et al. Increased susceptibility to ultraviolet-B and carcinogens of mice lacking the DNA excision repair gene XPA. *Nature* **377**, 169–173 (1995).
41. Cheo, D. L. et al. Characterization of defective nucleotide excision repair in XPC mutant mice. *Mutat. Res.* **374**, 1–9 (1997).
42. Liu, L. et al. Essential roles of S-nitrosothiols in vascular homeostasis and endotoxic shock. *Cell* **116**, 617–628 (2004).
43. Donehower, L. A. et al. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
44. Guyenet, S. J. et al. A simple composite phenotype scoring system for evaluating mouse models of cerebellar ataxia. *J. Vis. Exp.* **39**, 1787 (2010).
45. Dingler, F. A. et al. Two aldehyde clearance systems are essential to prevent lethal formaldehyde accumulation in mice and humans. *Mol. Cell* **80**, 996–1012.e9 (2020).
46. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet based single cell RNA sequencing data. *Gigascience* **9**, giaa151 (2020).
47. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019).
48. Popescu, D. M. et al. Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
49. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
50. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

51. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
52. Park, J. et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* **360**, 758–763 (2018).
53. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
54. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).

Acknowledgements We thank members of the K.J.P., G.P.C. and J.I.G. laboratories for critical reading of the manuscript. We thank J. Svejstrup and A. Tufegdzic Vidakovic for the pGEX-Dsk2 plasmid, E. Friedberg, G. T. van der Horst and J. Hoeijmakers for sharing the *Xpc*$^{-/-}$, *Xpa*$^{-/-}$ and *Csb*$^{m/m}$ mice and S. Wells, M. Stewart and H. Cater for work done at MRC Harwell. We thank Pfizer for the GDF15 monoclonal antibody. K.J.P. is supported by MRC, CRUK (C42693/A23273), Wellcome Trust (106202/Z/14/Z) and European Union Research and Innovation programme Horizon 2020 (Grant Agreement Number 730879). L.M., F.A.D. and G.P.C. were supported by CRUK (C42693/A23273). J.I.G. is supported by the Hubrecht Institute. C.L.M. is supported by the Wellcome Trust (106202/Z/14/Z). M.R.C. is funded by an NIHR Research Professorship (RP-2017-08-ST2-002). J.R.F. is funded by NIHR Cambridge Blood and Transplant Research Unit Organ Donation. Z.K.T. and M.R.C. are funded by MRC HCA grant MR/S035842/1. L.G. is funded from CRUK (FC0001166), the UK Medical Research Council (FC001166) and the Wellcome Trust (FC001166). S.O. is supported by the Medical Research Council MRC.MC.UU.12012.1, a Wellcome Senior Investigator Award 214274/Z/18/Z and the NIHR Cambridge Biomedical Research Centre. J.A.T. is supported by an NIHR Clinical Lectureship (CL-2019-14-504).

Author contributions K.J.P., J.I.G. and G.P.C. conceived the study. K.J.P. wrote the manuscript. L.M. and J.I.G. designed and performed the majority of the experiments. F.A.D. performed GDF15 measurements. M.R.C. and J.R.F. designed the scRNA-seq experiment, J.R.F. prepared tissues for sequencing and processed the scRNA-seq data. Z.K.T. analysed the scRNA-seq data. C.L.M. measured formaldehyde adducts. L.G. generated CSB KO HEK 293 cells. J.A.T. performed in situ hybridization experiments. M.J.A. characterized the kidney pathology. S.O. provided crucial insight on the study.

Competing interests The authors declare no competing interests.

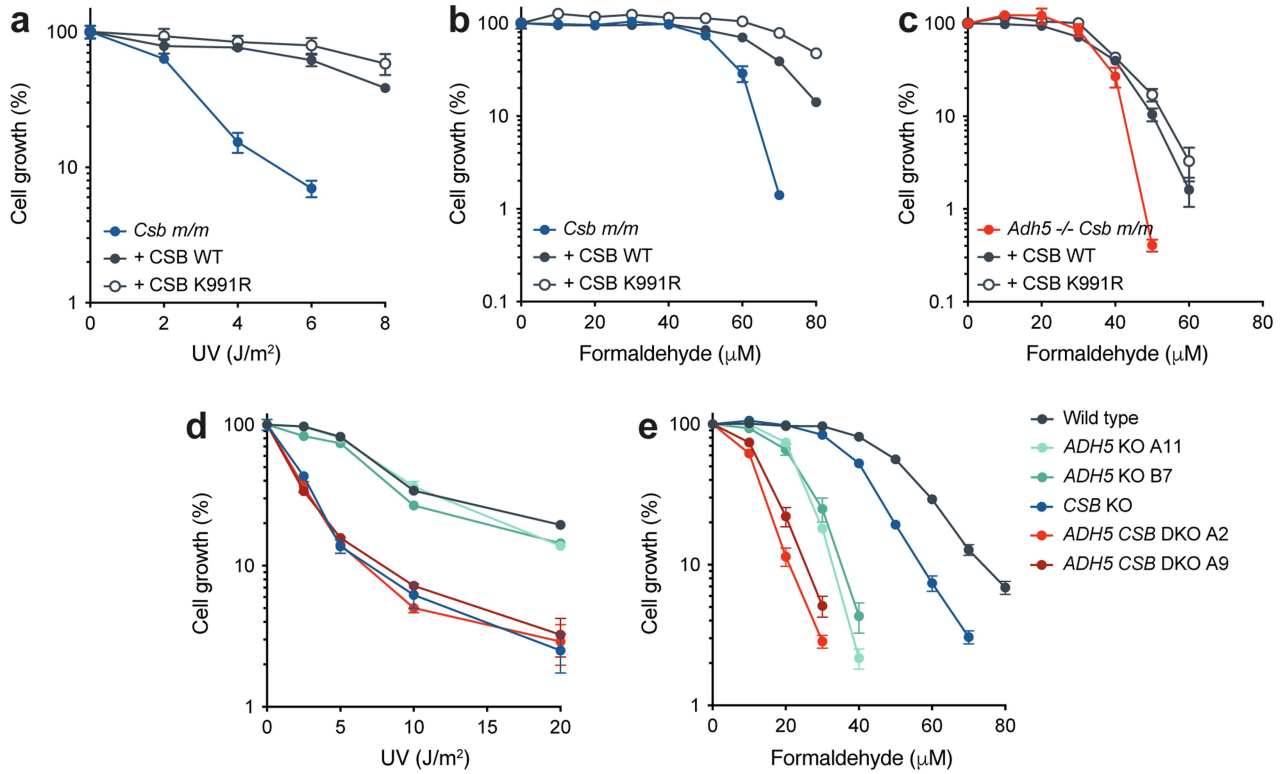
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-04133-7.

Correspondence and requests for materials should be addressed to Ketan J. Patel.

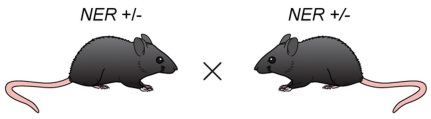
Peer review information *Nature* thanks Stephan Herzig, Ruben van Boxtel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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



Extended Data Fig. 1 | TC-NER and formaldehyde detoxification cooperate to protect cells from formaldehyde toxicity. a-c, Cytotoxicity of UV (a) and formaldehyde (b) in $Csb^{m/m}$ (a,b) or $Adh5^{-/-}Csb^{m/m}$ (c) tMEF cell lines complemented with either WT or K991R CSB. Data plotted as mean and s.e.m; experiments performed in triplicate. **d-e**, Cytotoxicity of UV and formaldehyde in HEK293 cell lines, data shown as mean and s.e.m., experiments performed in triplicate.

a

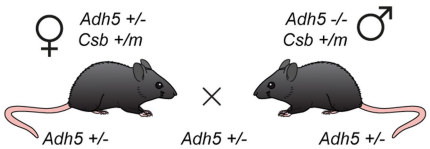


	NER +/+	NER +/-	NER -/-	
Exp.	25 %	50 %	25 %	
Obs. *Csb*	19.6 % n = 22 / 112	60.7 % n = 68 / 112	19.6 % n = 22 / 112	*P* = 0.3557
Obs. *Xpc*	26.7 % n = 35 / 131	54.2 % n = 71 / 131	19.1 % n = 25 / 131	*P* = 0.2339
Obs. *Xpa*	31.2 % n = 40 / 128	47.7 % n = 61 / 128	21.1 % n = 27 / 128	*P* = 0.4581

b

	Adh5 -/- NER +/+	Adh5 -/- NER +/-	Adh5 -/- NER -/-	
Exp.	25 %	50 %	25 %	
Obs. *Csb*	36.5 % n = 61 / 167	58.0 % n = 97 / 167	5.4 % n = 9 / 167	*P* < 0.0001
Obs. *Xpc*	20.1 % n = 9 / 44	59.1 % n = 26 / 44	20.1 % n = 9 / 44	*P* = 0.6109
Obs. *Xpa*	34.6 % n = 37 / 107	57.9 % n = 62 / 107	7.5 % n = 8 / 107	*P* = 0.0004

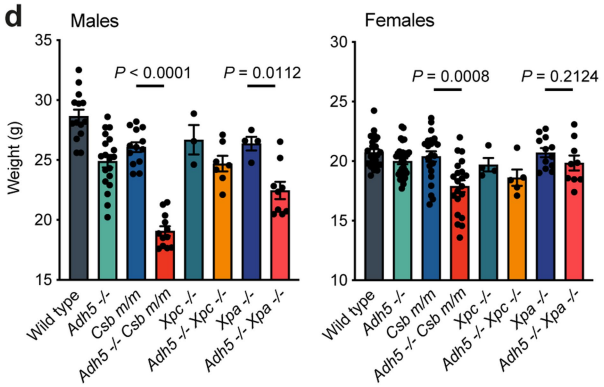
c



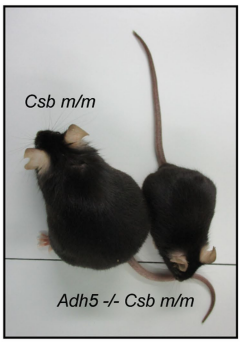
	Adh5 +/- Csb +/+	Adh5 +/- Csb +/m	Adh5 +/- Csb m/m	
Exp.	12.5 %	25 %	12.5 %	
Obs.	16.3 % n = 24 / 147	21.8 % n = 32 / 147	10.9 % n = 16 / 147	

	Adh5 -/- Csb +/+	Adh5 -/- Csb +/m	Adh5 -/- Csb m/m	
Exp.	12.5 %	25 %	12.5 %	
Obs.	10.2 % n = 15 / 147	33.3 % n = 49 / 147	7.5 % n = 11 / 147	*P* = 0.1710

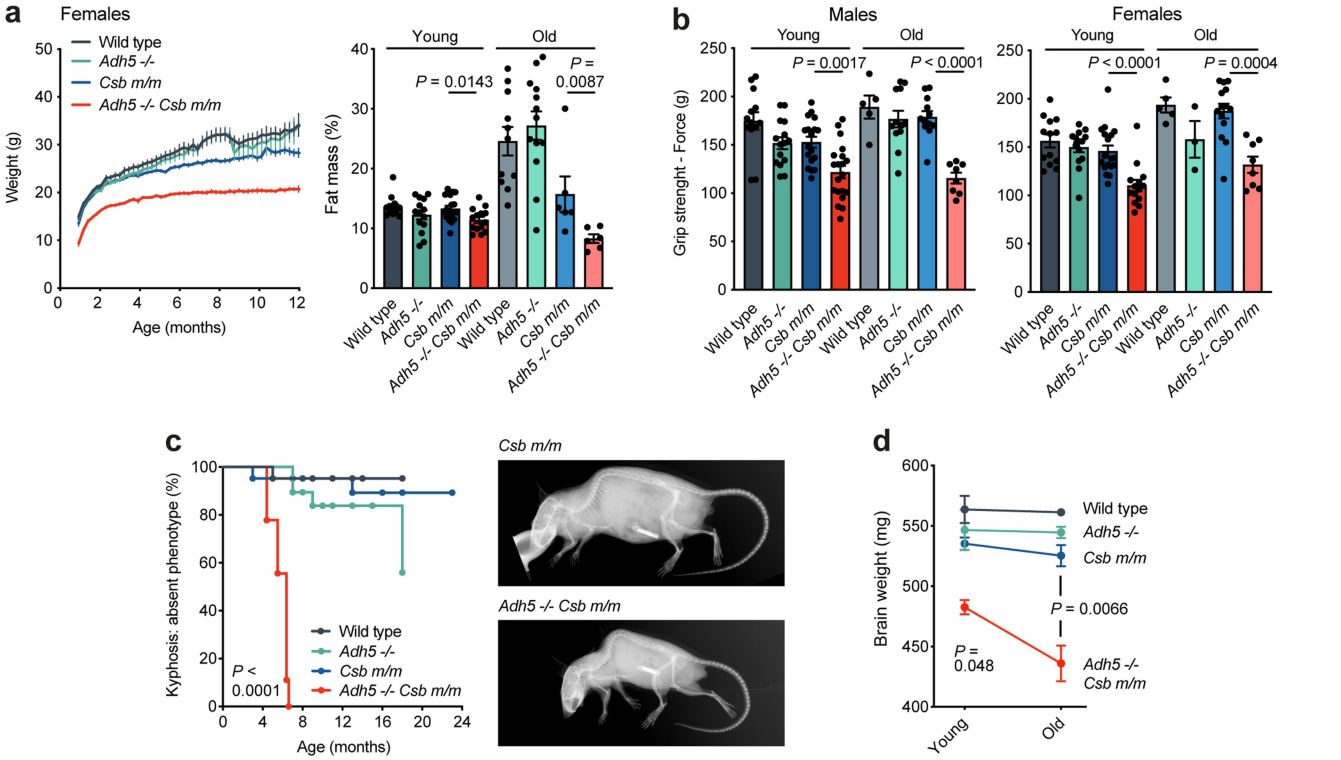
d



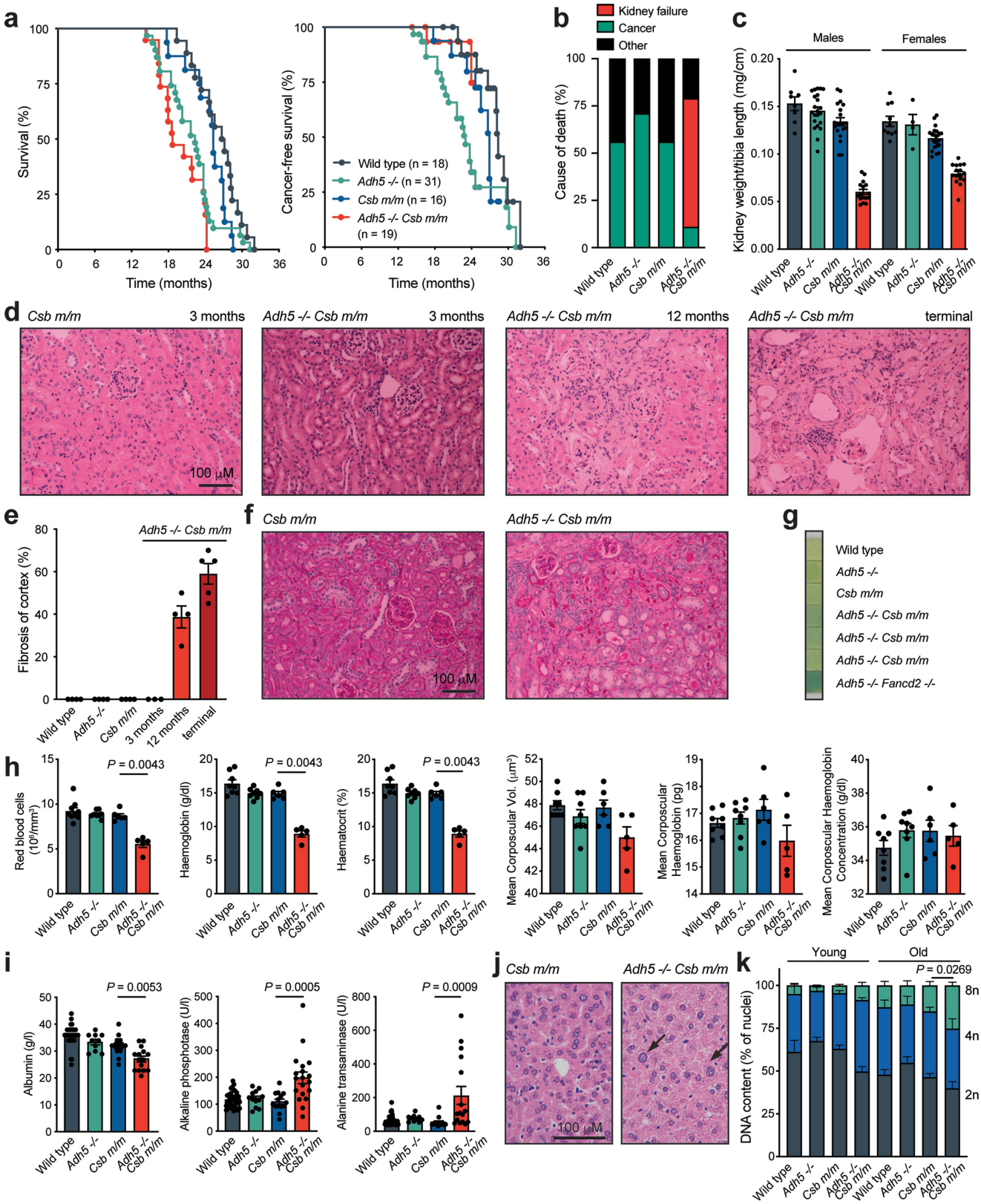
e



Extended Data Fig. 2 | *Adh5*$^{-/-}$ *Csb*$^{m/m}$ mice are born at sub-Mendelian ratios and have reduced body weight. a, Ratios of pups from *Csb*$^{+/m}$, *Xpc*$^{+/-}$ and *Xpa*$^{+/-}$ crosses showing that homozygous mice are born at Mendelian ratios (*P* calculated by two-sided Chi-squared test) **b**, Ratios of pups born from *Adh5*$^{-/-}$*Csb*$^{+/m}$, *Adh5*$^{-/-}$*Xpc*$^{+/-}$ and *Adh5*$^{-/-}$*Xpa*$^{+/-}$ crosses showing that both *Adh5*$^{-/-}$*Csb*$^{m/m}$ *and Adh5*$^{-/-}$*Xpa*$^{-/-}$ pups are born at sub-Mendelian ratios (*P* calculated by two-sided Chi-squared test). **c**, Ratios of pups born from *Adh5*$^{+/-}$ *Csb*$^{+/m}$ and *Adh5*$^{-/-}$*Csb*$^{+/m}$ crosses showing the ratio of *Adh5*$^{-/-}$*Csb*$^{m/m}$ pups is partially rescued when the mother is aldehyde-detoxification proficient (*Adh5* +/- instead of *Adh5* -/- in **b**, *P* calculated by two-sided Chi-squared test). For **a, b**, and **c**, mice were genotyped between 2-3 wk of age. **d**, Weights of adult male and female mice at 8 wk of age (Data shown as mean and s.e.m.; *P* calculated by two-sided Mann-Whitney test; *n* = 14, 19, 13, 12, 4, 9, 3, 7 for males left to right and 24, 30, 22, 21, 12, 9, 4, 5 for females left to right). **e**, Image of *Csb*$^{m/m}$ and *Adh5*$^{-/-}$*Csb*$^{m/m}$ littermates at 12 months of age.

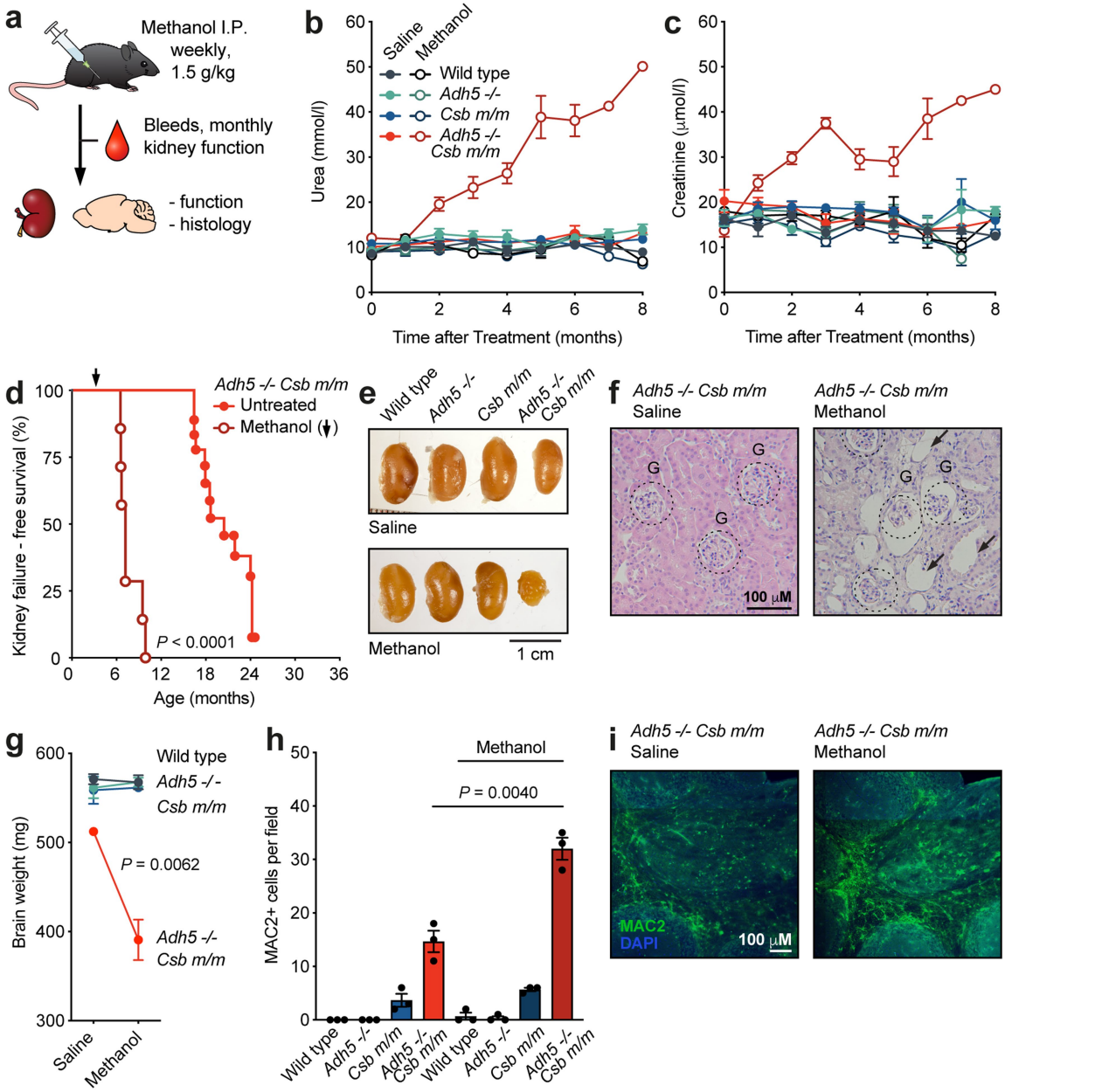


Extended Data Fig. 3 | *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice exhibit features of human Cockayne Syndrome. a, Growth curves of female mice based on weekly weights. Data shown as mean and s.e.m. $n = 14, 14, 18$ and 14. Along with fat mass from EchoMRI performed at 11 wk (young) and 53 wk (old). *P* calculated by two-sided Mann-Whitney test; data shown as mean and s.e.m.; $n = 13, 14, 18, 14, 11, 13, 6$ and 6 left to right. **b**, Bar graph of grip strength for young (3 months) and old (1 yr) mice, determined by placing all four limbs on a grid attached to a force gauge. *P* calculated by two-sided Mann-Whitney test, data shown as mean and s.e.m.; $n = 13, 15, 18, 20, 5, 12, 12$ and 8 for males left to right and 13, 14, 18, 14, 5, 3, 14 and 8 for females left to right. **c**, Left, age of onset plot for kyphosis. *P* calculated by two-sided Mantel-Cox logrank test, $n = 21, 19, 22$ and 9. Right, representative x-rays of *Csb*$^{m/m}$ and *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice at 1 yr of age showing kyphosis in the *Adh5*$^{-/-}$*Csb*$^{m/m}$ mouse. **d**, Brain weights of mice taken at 3 months (young) and 18 months (old). *P* calculated by two-tailed Student's *t*-test; data shown as mean and s.e.m., $n = 3$ mice.

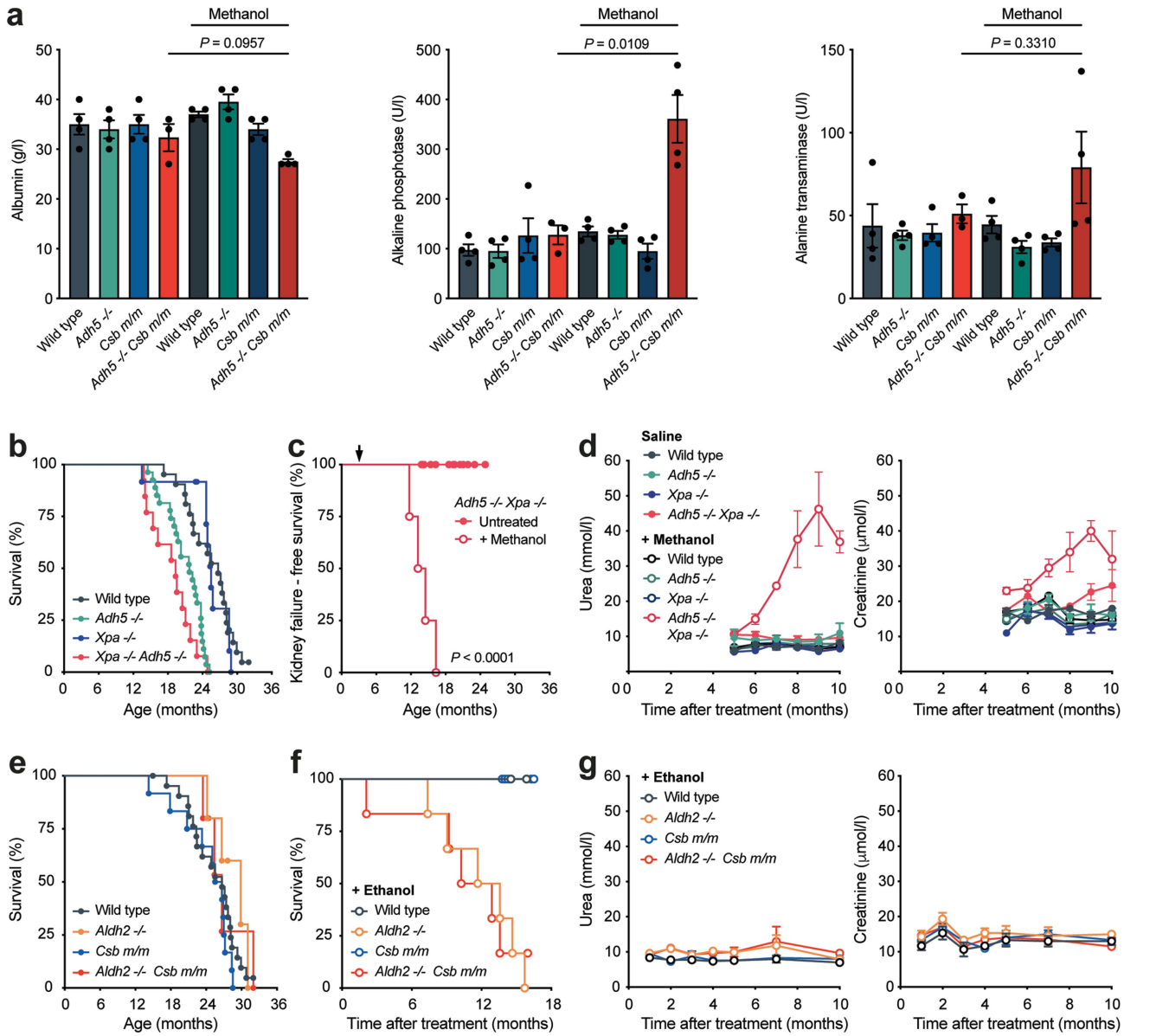


Extended Data Fig. 4 | See next page for caption.

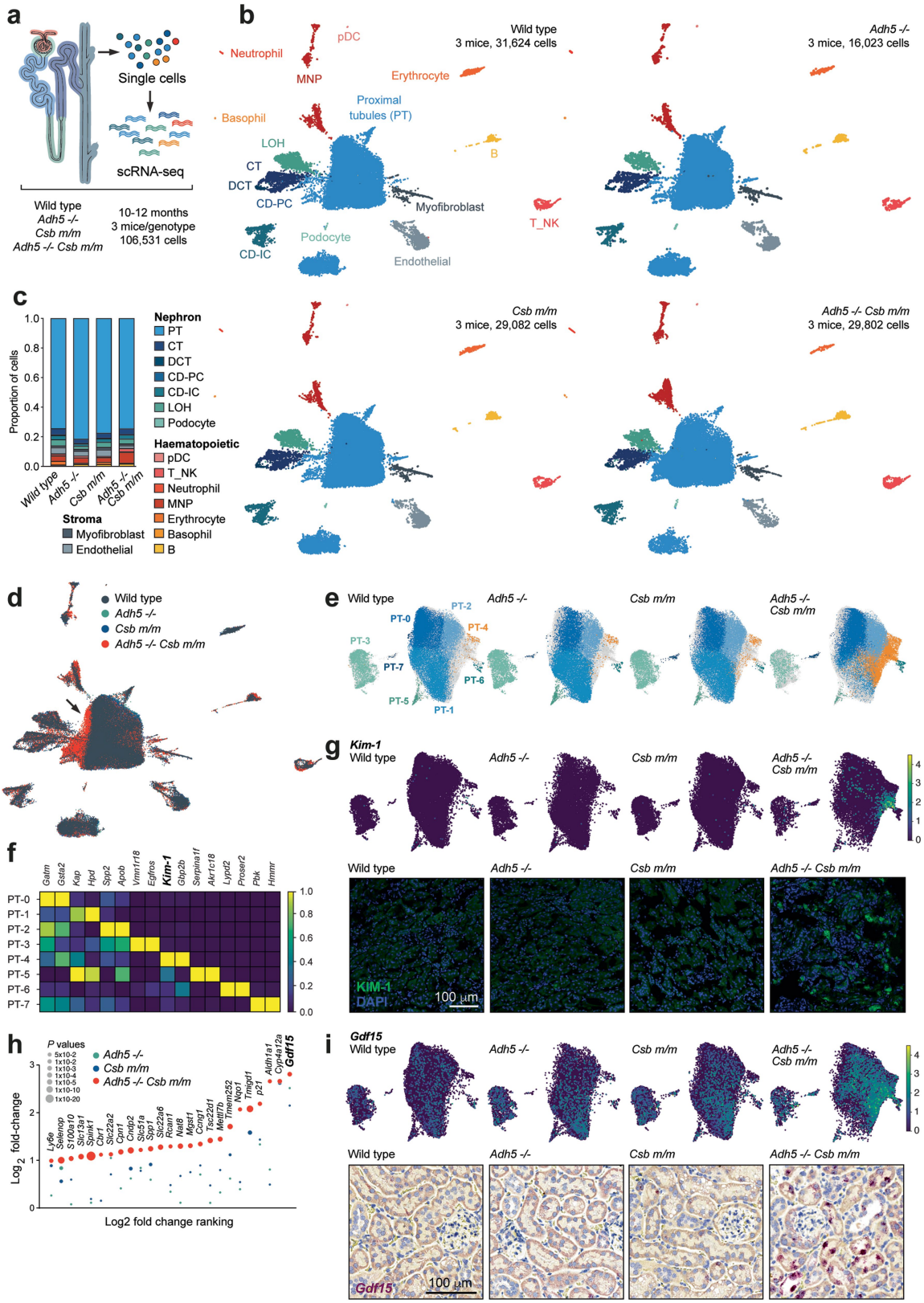
Extended Data Fig. 4 | *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice succumb to chronic kidney failure and have liver abnormalities. a, Survival and cancer-free survival curve of *Adh5*$^{-/-}$*Csb*$^{m/m}$ and control mice. **b**, Bar chart indicating the cause of death for wild type, *Adh5*$^{-/-}$, *Csb*$^{m/m}$ and *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice (n = 18, 31, 16 and 19). **c**, Weights of *Adh5*$^{-/-}$*Csb*$^{m/m}$ male and female kidneys at 12 months relative to tibia length, data shown as mean and s.e.m. (n=8, 22, 20, 16, 10, 4, 22, 17 from left to right). **d**, Representative H&E stained sections of kidney from *Csb*$^{m/m}$ and *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice at sequential timepoints (n=4, 3, 4, 5, from left to right). **e**, Bar chart of the percentage of fibrosis in the cortex of H&E stained kidney sections at sequential time points in *Adh5*$^{-/-}$*Csb*$^{m/m}$ and terminal controls, data shown as mean and s.e.m. (n=4, 4, 4, 3, 4, 5, from left to right). **f**, Representative PAS stained sections of kidney showing intratubular casts from *Csb*$^{m/m}$ and *Adh5*$^{-/-}$*Csb*$^{m/m}$ terminal mice (n=3). **g**, Urine obtained from indicated mice were tested for the presence of proteinuria by multistix 10SG. **h**, Blood counts from terminal blood samples of *Adh5*$^{-/-}$*Csb*$^{m/m}$ and controls P calculated by two-sided Mann-Whitney test; data show as mean and s.e.m.; n = 9, 8, 6, 5 from left to right. **i**, Terminal serum measurements of albumin, alkaline phosphatase and alanine transaminase. P calculated by two-sided Mann-Whitney test; data show as mean and s.e.m.; n = 33, 10, 13 and 16. **j**, Representative H&E stained sections of liver from age-matched *Csb*$^{m/m}$ and *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice, arrows indicate cells with enlarged nuclei (n=3). **k**, Quantification of hepatocyte nuclear DNA content in young (3 month) and old (18 month) mice. P calculated by two-sided Student's t-test for the content of 8n nuclei; data shown as mean and s.e.m.; n = 3 mice.



Extended Data Fig. 5 | Methanol exposure exacerbates the Cockayne Syndrome phenotype in *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice. a, Scheme outlining weekly intra-peritoneal (I.P.) injection of 1.5 g/kg methanol (or saline) and analysis of treated mice. **b**-**c**, Monthly serum levels of urea (b) and creatinine (c) from mice exposed to methanol and saline controls, data plotted as mean and s.e.m.; $n = 4$. **d**, Kidney failure-free survival curve of *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice with and without methanol exposure (*P* calculated by two-sided Mantel-Cox logrank test; $n = 7$ and 19). **e**, Representative image of kidneys from age-matched mice exposed to methanol or saline, taken after 24 hr of fixation. **f**, Representative H&E stained sections of kidney from *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice exposed to methanol or saline, G indicates glomeruli and arrows indicate atrophic tubules *(n=4)*. **g**, Brain weights of age-matched mice exposed to methanol or saline. *P* calculated by two tailed Student's *t*-test; data shown as mean and s.e.m., $n = 3$ mice. **h**, Quantification of the number of MAC2+ cells per field in mice exposed to methanol or saline. *P* calculated by two tailed Student's *t*-test; data shown as mean and s.e.m., $n = 3$ mice. **i**, Representative immunofluorescence images of the cerebellum of *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice exposed to methanol or saline stained with MAC2 and DAPI at 40x *(n=3)*.

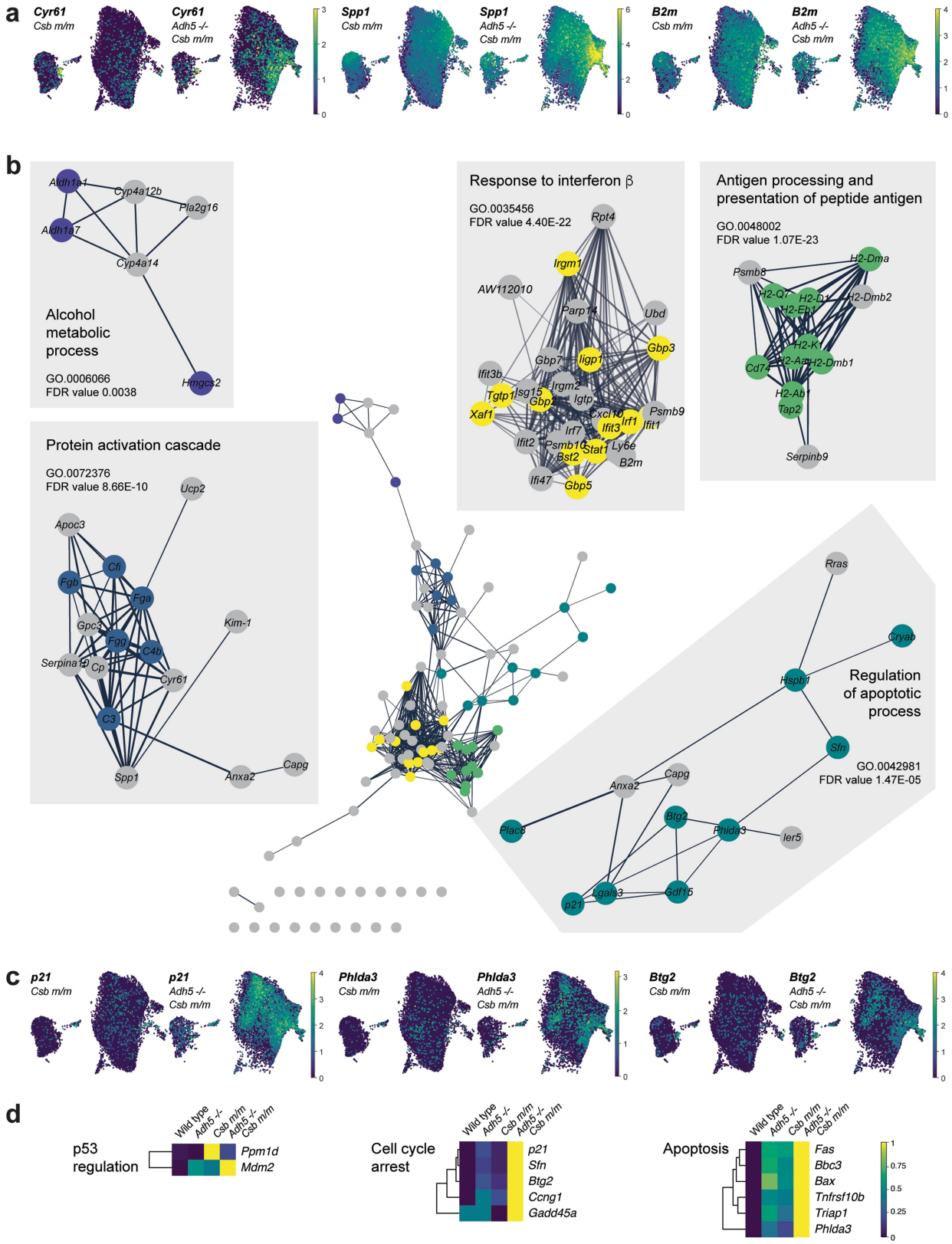


Extended Data Fig. 6 | Methanol reveals kidney failure in *Adh5*$^{-/-}$*Xpa*$^{-/-}$ mice but ethanol does not in *Aldh2*$^{-/-}$*Csb*$^{m/m}$ mice. a, Serum measurements of albumin, alkaline phosphatase and alanine transaminase after 6 months of methanol treatment along with saline-treated controls. *P* calculated by two-sided Mann-Whitney test; data show as mean and s.e.m.; $n = 4$ mice. **b**, Kaplan-Meier survival curve of wild type, *Adh5*$^{-/-}$, *Xpa*$^{-/-}$ and *Adh5*$^{-/-}$*Xpa*$^{-/-}$ mice ($n = 18, 31, 10$ and 13). **c**, Kidney failure-free survival curve of *Adh5*$^{-/-}$*Xpa*$^{-/-}$ mice with and without methanol exposure (*P* calculated by two-sided Mantel-Cox logrank test; $n = 4$ and 13). **d**, Monthly serum levels of urea and creatinine from mice exposed to methanol or saline, data plotted as mean and s.e.m.; n = 4. **e**, Kaplan-Meier survival curve of wild-type, *Aldh2*$^{-/-}$, *Csb*$^{m/m}$ and *Aldh2*$^{-/-}$*Csb*$^{m/m}$ mice ($n = 18, 5, 16$ and 5). **f**, Kaplan-Meier survival curve of wild-type, *Aldh2*$^{-/-}$, *Csb*$^{m/m}$ and *Aldh2*$^{-/-}$*Csb*$^{m/m}$ mice treated with 20% ethanol continuously in the drinking water ($n = 6$). **g**, Monthly serum levels of urea and creatinine from mice exposed to 20% ethanol continuously in the drinking water, data plotted as mean and s.e.m.; $n = 6$.

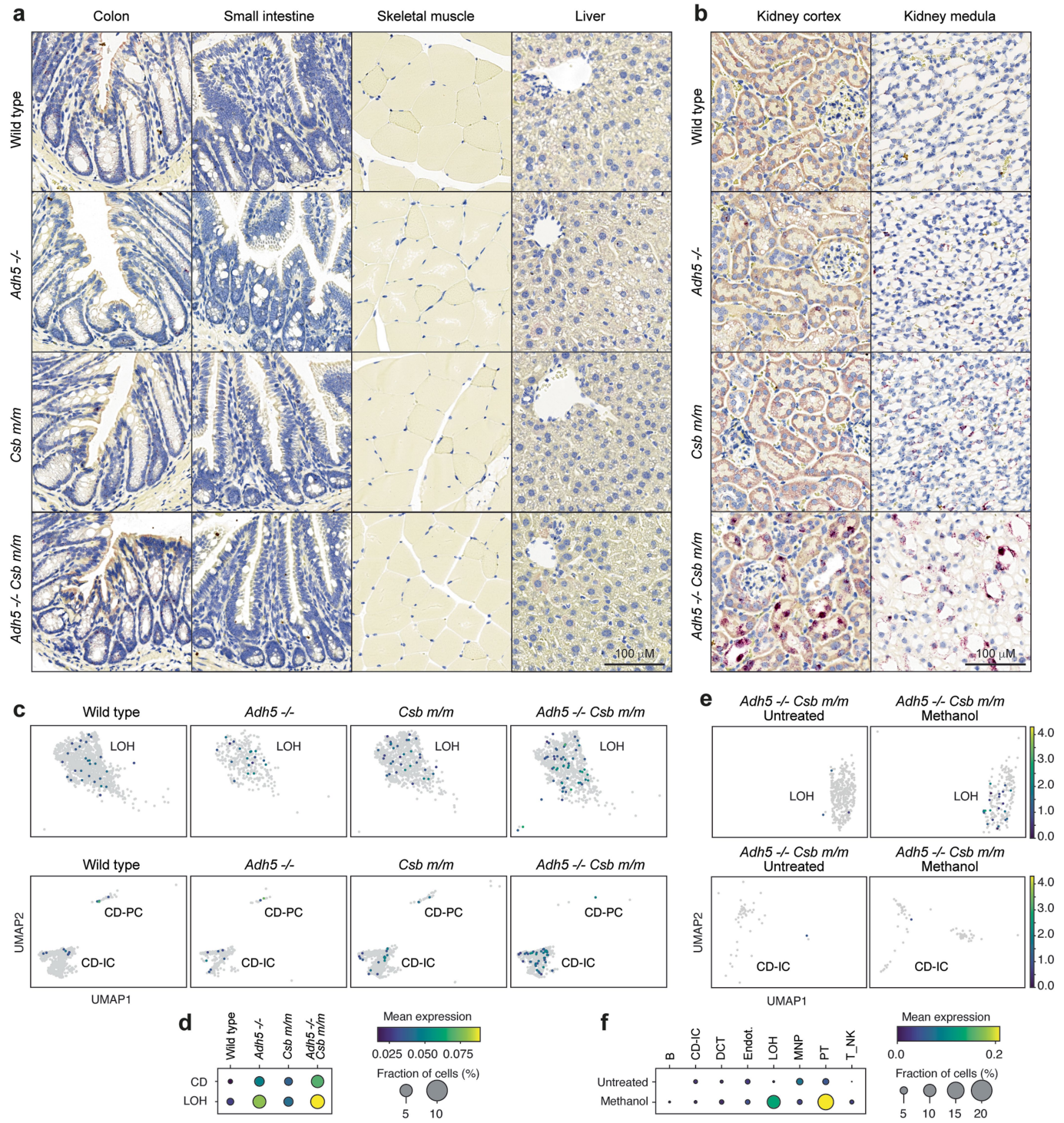


Extended Data Fig. 7 | See next page for caption.

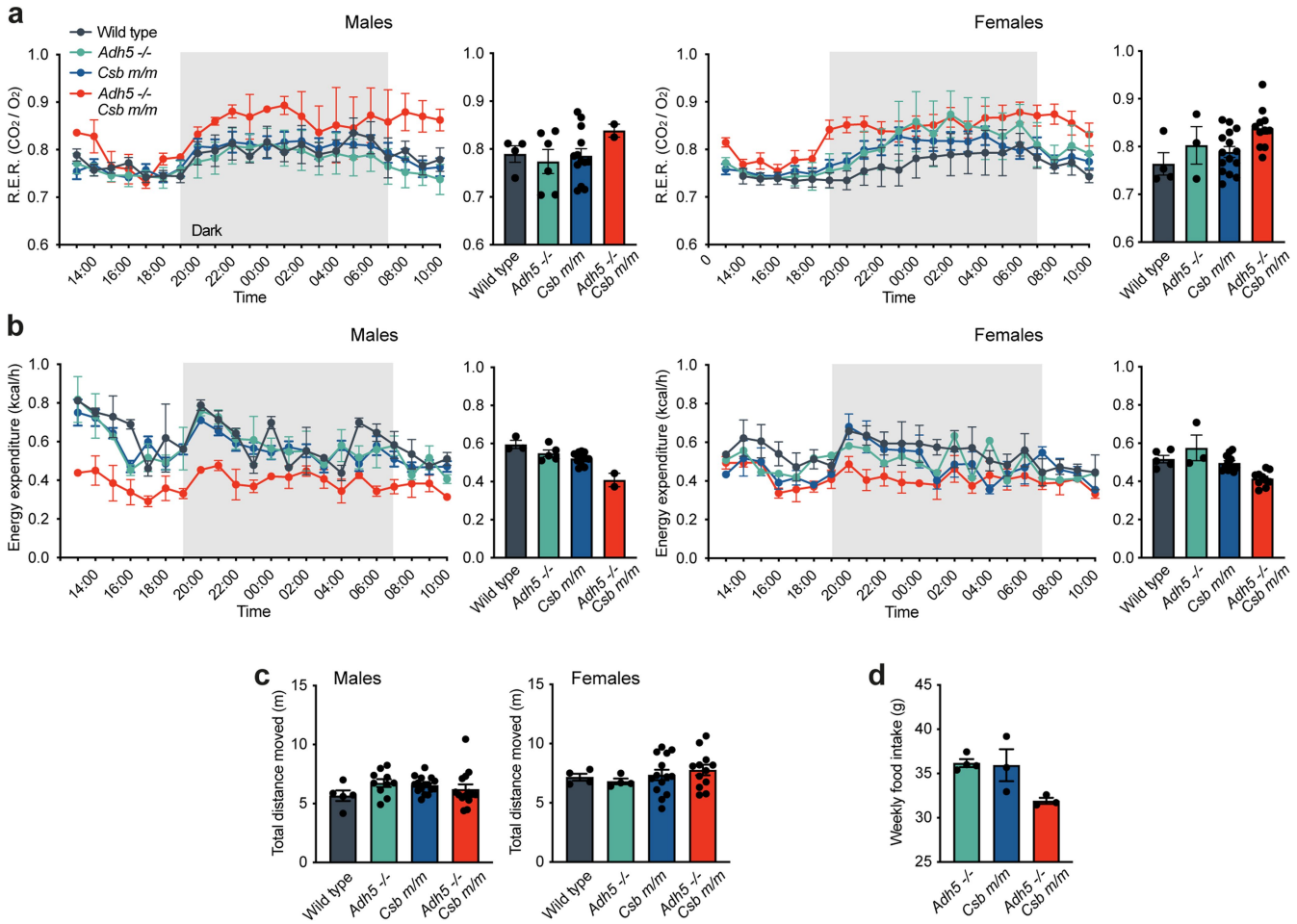
Extended Data Fig. 7 | scRNA-seq reveals cells susceptible to formaldehyde transcriptional stress. a, Scheme outlining the scRNA-seq experiment. **b**, UMAP plots of murine kidney scRNA-seq data for wild type, *Adh5*$^{-/-}$, *Csb*$^{m/m}$ and *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice, $n = 3$ and $n = 31{,}624, 16{,}023, 29{,}082, 29{,}802$ cells (pDC, plasmacytoid dendritic cell; MNP, mononuclear phagocyte; LOH, loop of henle; CT, connecting tubule; DCT, distal convoluted tubules; CD-PC, collecting duct – principle cell; CD-IC, collecting duct – intercalated cell; T_NK, T cells/NK cells; B, B cells). **c**, Bar chart showing the composition of cell types in the scRNA-seq data. **d**, UMAP plot of all scRNA-seq data with each genotype labelled a different colour, arrow indicates PT cells that are distinct to *Adh5*$^{-/-}$*Csb*$^{m/m}$ kidneys. **e**, UMAP plot of PT sub-clusters split by genotype. **f**, Heatmap indicating the top two marker genes for each PT sub-cluster (marker genes are calculated by comparing the expression of each PT sub-cluster against all remaining PTs). **g, top**, Feature plots of *Kim-1* expression from PT sub-clusters of *Adh5*$^{-/-}$*Csb*$^{m/m}$ cells and controls, **bottom**, immunofluorescence images of *Adh5*$^{-/-}$*Csb*$^{m/m}$ and control kidney sections stained with Kim-1 taken from mice aged 12 months. **h**, Log2 fold change plot of DEGs from PT-4 made relative to wild-type PT-4 cells. **i**, **top**, Feature plots of *Gdf15* expression from PT sub-clusters of *Adh5*$^{-/-}$*Csb*$^{m/m}$ cells and controls, **bottom**, representative images of *In situ* hybridisation for Gdf15 mRNA (red spots) performed on *Adh5*$^{-/-}$*Csb*$^{m/m}$ and control kidney sections, n=3.



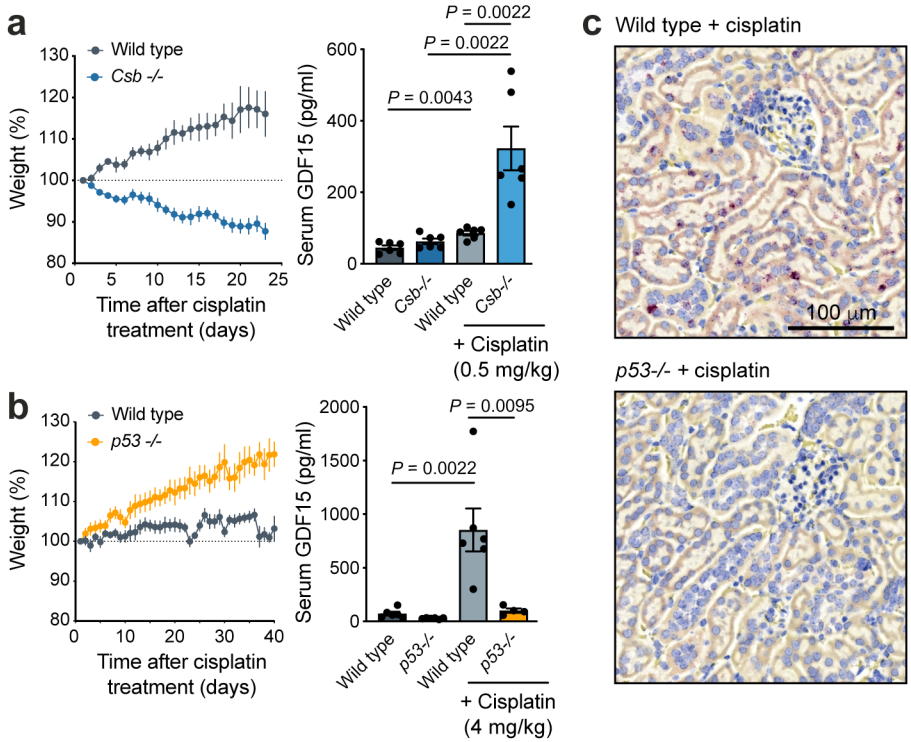
Extended Data Fig. 8 | scRNA-seq gene expression analysis of PT-4 sub-cluster cells. a, Feature plots of *Cyr61, Spp1* and *B2m* expression from PT sub-clusters of $Csb^{m/m}$ and $Adh5^{-/-}Csb^{m/m}$ cells. **b**, Network of top 100 marker genes from PT sub-cluster PT-4 (marker genes were calculated by comparing expression of genes in PT-4 to all other PT cells). Data visualised in cytoscape v3.7.1 using the STRING app. Nodes are coloured based on top GO term enrichment pathway. **c**, Feature plots of *p21, Phlda3* and *Btg2* expression from PT sub-clusters of $Csb^{m/m}$ and $Adh5^{-/-}Csb^{m/m}$ cells. **d**, Heatmap of expression of p53 target genes in PT cells of $Adh5^{-/-}Csb^{m/m}$ and control mice.



Extended Data Fig. 9 | *Gdf15* expression in mouse tissues. a and b, Representative images of *in situ* hybridisation for *Gdf15* mRNA (red spots) performed on *Adh5*$^{-/-}$*Csb*$^{m/m}$ and control kidney sections, *n*=3 **c**, Feature plots of *Gdf15* expression from LOH and CD cells for *Adh5*$^{-/-}$*Csb*$^{m/m}$ and control mice. **d**, Dot plot of mean expression and fraction of cells expressing *Gdf15* in the loop of henle (LOH) and collecting duct (CD) for *Adh5*$^{-/-}$*Csb*$^{m/m}$ and control mice. **e**, Feature plots of *Gdf15* expression from LOH and CD cells for *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice treated with methanol or untreated. **f**, Dot plot of mean expression and fraction of cells expressing *Gdf15* in *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice treated with methanol or untreated.



Extended Data Fig. 10 | Metabolic analysis of *Adh5*$^{-/-}$*Csb*$^{m/m}$ mice. a-b, Male and female mice were placed in metabolic chambers and respiratory exchange rate (R.E.R) and energy expenditure were measured over a 20-hr period, data shown as mean and s.e.m. $n = 3, 5, 13$ and 2 for males and $n = 7, 4, 11$ and 10 for females. **c**, male and female mice were placed in open field chambers and activity was measured over 20 min, data shown as mean and s.e.m (n=5, 10, 16 and 14 for males and n= 4, 4, 14 and 12 for females). **d**, Bar chart of weekly food intake from *Adh5*$^{-/-}$*Csb*$^{m/m}$ and control mice singly housed and averaged out over 5 wk, data shown as mean and s.e.m, n=3.



Extended Data Fig. 11 | GDF15 release in response to cisplatin is increased in *Csb*$^{m/m}$ mice but absent in *p53*$^{-/-}$ mice. a, Daily weights of wild-type and *Csb*$^{m/m}$ mice exposed to weekly 0.5 mg/kg cisplatin intra-peritoneal injections alongside serum GDF15 measurements taken before and 4 wk after weekly injections. *P* calculated using two-sided Mann-Whitney test; data shown as mean and s.e.m.; *n* = 6 mice. **b**, Daily weights of wild-type and *p53*$^{-/-}$ mice exposed to weekly 4 mg/kg cisplatin intra-peritoneal injections alongside serum GDF15 measurements taken before and 6 wk after weekly injections. *P* calculated using two-sided Mann-Whitney test; data shown as mean and s.e.m.; *n* = 6 mice. **c**, *In situ* hybridisation for *Gdf15* mRNA (red spots) performed on wild-type and *p53*$^{-/-}$ kidney sections 24 h after 4 mg/kg cisplatin treatment (*n*=3).

nature research

Corresponding author(s): KJ Patel

Last updated by author(s): 23/06/21

Reporting Summary

Nature Research 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 Research 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	No commercial, open source or custom code was used for data collection in this study
Data analysis	For scRNA-seq data analysis: Following sequencing, BCL files were demutiplexed to Fastq files using CASAVA. Subsequently splitting to single cells and mapping of genes was carried out using Cellranger v3.0. The single-cell data (10x cellranger output) was corrected for ambient RNA expression using SoupX (v1.2.1)51. After SoupX, doublet detection was performed using scrublet (v0.2.1)52 with adaptations outlined in Popescu et al53. Briefly, after scrublet was performed, the data was iteratively sub-clustered using standard Seurat-inspired scanpy (v.1.4.5.post2) workflow 54,55 and a median scrublet score for each subcluster was computed. Median absolute deviation was computed from the cluster scrublet scores and a one tailed t-test was performed with Benjamini-Hochberg (BH) correction 56 applied where cells with significantly outlying cluster scrublet scores (BH pval < 0.1) were flagged as potential doublets. The data was then processed using scanpy with standard quality control steps; cells were filtered if number of genes > 2500 or < 200. Similar to Park et al 57, percentage mitochondrial content cut-off was set at <50%. Genes were retained if expressed by at least 3 cells. Gene counts for each cell were normalised to contain a total count equal to 10000 per cell. Highly variable genes were selected based on the following parameters: minimum and maximum mean expression are >=0.0125 and <=3

respectively; minimum dispersion of genes = 0.5. The effects of total counts per cell, percentage of mitochondrial content and cell cycle states were regressed out prior to scaling of variable genes. Neighbourhood graph graphs were constructed with 20 neighbours using 80 PCs (all cells) or 30 PCs (proximal tubules only). Clustering was performed using Leiden algorithm58 with resolution set at 0.5. Uniform Manifold Approximation and Projection (UMAP; v3.10.0)59 was used for dimensional reduction and visualization; Partition-based graph abstraction (PAGA)-initialization was used to compute the UMAP embedding for all cells while default (spectral) was used for PT; the minimum distance was set at 0.3; all other parameters were used as per default settings in scanpy.

Differential gene testing was performed using the Wilcoxon test rank sum implemented in scanpy's rank_genes_groups module. BH post-hoc correction was applied to the p-values calculated after DEG analysis. Genes that attained corrected P value of < 0.05 were considered statistically significant.

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All single cell RNA-sequencing data has been deposited on geo with reference GSE175792 ready to go live on the day of publication.

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	For tissue culture work all experiments were repeated 3 times and averages plotted. This was predetermined as a matter of good practise rather than predetermined by statistical methods. These sample sizes allowed for us to detect statistically significant differences and so further experiments were not needed. For the majority of the experiments performed on live mice the sample size was predetermined at 10 so that we could generate robust statistical significant data, however we have included additional animals depending on the genotyping from breeding crosses set up to provide 10 mice per genotype. For more invasive studies such as the methanol, cisplatin, GDF15 Ab studies we designed experiments to include at least n= 3 for robust statistics. The majority of these experiments were repeated so that in most cases n=6. The data generated from these experiments resulted in statistical significance therefore more experiments were not needed. These numbers were based on the likely number of mice for each genotype that we would be able to generate from our capacity of mouse cages, we also wanted to refrain from using too many mice unnecessarily if statistical significance could already be reached. For immunoflourescence analysis/histology and RNAscope sample sizes were equal to 3 mice, this was predetermined as a matter of good practise and also allowed for robust findings so that experiments did not need repeating. For single cell analysis we used 3 mice per genotype and pooled the samples for sequencing this was also predetermined to average out any effects and ensure robust findings. When analysing the data we were able to make statistcal significant observations between genotypes so we did not need to repeat the experiment.
Data exclusions	No data was excluded from the analysis
Replication	For tissue culture work all experiments were repeated 3 times. For animal studies where mice were given methanol or cisplatin all experiments were repeated at least once and all data plotted.

For live animal experiments other than cisplatin and methanol exposure protocols animals were divided up into cohorts as they were generated and phenotyped at predetermined time points. In total 6 cohorts were generated and data pooled as it was generated.

All attempts at replication were successful.

Randomization

For live animals, once genotypes were confirmed cohorts were organised and placed into experimental protocols. Cohorts were designed so that they had at least 3 animals per genotype.

For tissue culture experiments experiments were designed so that all cell lines were tested in a single experiment, randomisation was not required as all lines were treated the same.

Blinding

For Tissue culture experiments blinding was not deemed necessary as all groups were treated the same and experiments were performed in triplicate.

For all mouse work technicians were blinded to the genotype of the animals and mouse I.D numbers were used.

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

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

For screening HEK293 cells an in-house rabbit anti-ADH5 antibody was used which has been previously described. For screening complemented tMEF clones anti-FLAG (Sigma, clone M2) was used at 1/1000. For the degradation of RPB1 blots, RPB1 (Cell Signalling, D8L4Y) was used at 1/1000 and for ubiquitylation of RPB1, RPB1 (abcam, 4H8) was used at 1/10,000. In both RPB1 blots, mouse anti-vinculin (Sigma, Cat V9131) was used as a loading control at 1/10,000. For immunoflourescence MAC2 (Cedarlane, clone M3/38) at 1/2000, Alexaflour 488 anti-rat (invitrogen) at 1/1000 and KIM-1 (R&D Systems, AF1817) used at 1/500. For GDF15 antibody treatment 10mg/kg was used, antibody provided by Pfizer (not commercial).

Validation

All antibodies used are commercially available with details above. Except the in house anti-Adh5 antibody, the details for this can be found here:
Pontel, L. B. et al. Endogenous formaldehyde is a hematopoietic stem cell genotoxin and metabolic carcinogen. Mol. Cell 60, 177–188 (2015).
The GDF15 antibody provided by Pfizer was previously characterised and details can be found here:
Breen, D. M. et al. GDF-15 Neutralization Alleviates Platinum-Based Chemotherapy-Induced Emesis, Anorexia, and Weight Loss in Mice and Nonhuman Primates. Cell Metab. 32, 938-950.e6 (2020).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

tMEF cell lines were made in house as described in the method section of the manuscript. HEK293t cells were a kind gift from the Svejstrup lab and were orginally purchased from Thermo Fisher Scientific.

Authentication

Cell lines were validated by Western blot as detailed in the methods

Mycoplasma contamination

All cell lines were routinely tested for mycoplasma contamination and were negative

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified lines were used

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	All lab animals used were in the C57 Black 6 background except p53 mice and wild type controls which were an F1 cross between C57 Black 6 and 129. Both males and females were used for in vivo phenotyping which was carried out at young (3months of age) and old time points (12months). For single cell analysis males aged between 10-12 months were used. Methanol exposure was done at 8-12 weeks on female mice. Cisplatin exposure was performed on males between 8-12 weeks old. GDF15 antibody experiments were performed on male mice between 12-20 weeks old. Animal experiments undertaken in this study were with approval of the MRC Laboratory of Molecular Biology animal welfare and ethical review body and the UK Home Office under the Animal (Scientific Procedures) Act 1986 license PFC07716E. All mice were maintained under specific pathogen-free conditions in individually ventilated cages (Techniplast GM500, Techniplast) on Ligno-cel FS14 spruce bedding (IPS, LTD) with environmental enrichment (fun tunnel, chew stick, and Enviro-Dri nesting material (LBS)) at 19–23°C with light from 7:00 am to 7:00 pm and fed Dietex CRM pellets (Special Diet Services) ad libitum.
Wild animals	No wild animals were used
Field-collected samples	No field collected samples were used
Ethics oversight	All animal work in this study were with approval from the MRC LMB animal welfare and ethical review body and the UK Home Office under the Animal (Scientific Procedures) Act 1986 license PFC07716E

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