Supplemental Material

Deficiency of nucleotide excision repair is associated with mutational signature observed in cancer

Myrthe Jager[1,7], Francis Blokzijl[1,2,7], Ewart Kuijk[1], Johanna Bertl[3,4], Maria Vougioukalaki[5], Roel Janssen[1], Nicolle Besselink[1], Sander Boymans[1], Joep de Ligt[1], Jakob Skou Pedersen[3], Jan Hoeijmakers[5], Joris Pothof[5], Ruben van Boxtel[1,6,*] and Edwin Cuppen[1,*].

[1]Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, Utrecht University, Universiteitsweg 100, 3584, CG, Utrecht, The Netherlands

[2]Present address: Oncode Institute, Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, 3584 CT Utrecht, The Netherlands.

[3]Department of Molecular Medicine, Aarhus University, Palle Juul-Jensens Boulevard 99, 8200 Aarhus N, Denmark

[4]Present address: Department of Mathematics, Aarhus University, Ny Munkegade 118, 8000 Aarhus C, Denmark

[5]Erasmus Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands

[6]Present address: Princess Máxima Center for Pediatric Oncology and Oncode Institute, 3584 CT Utrecht, The Netherlands

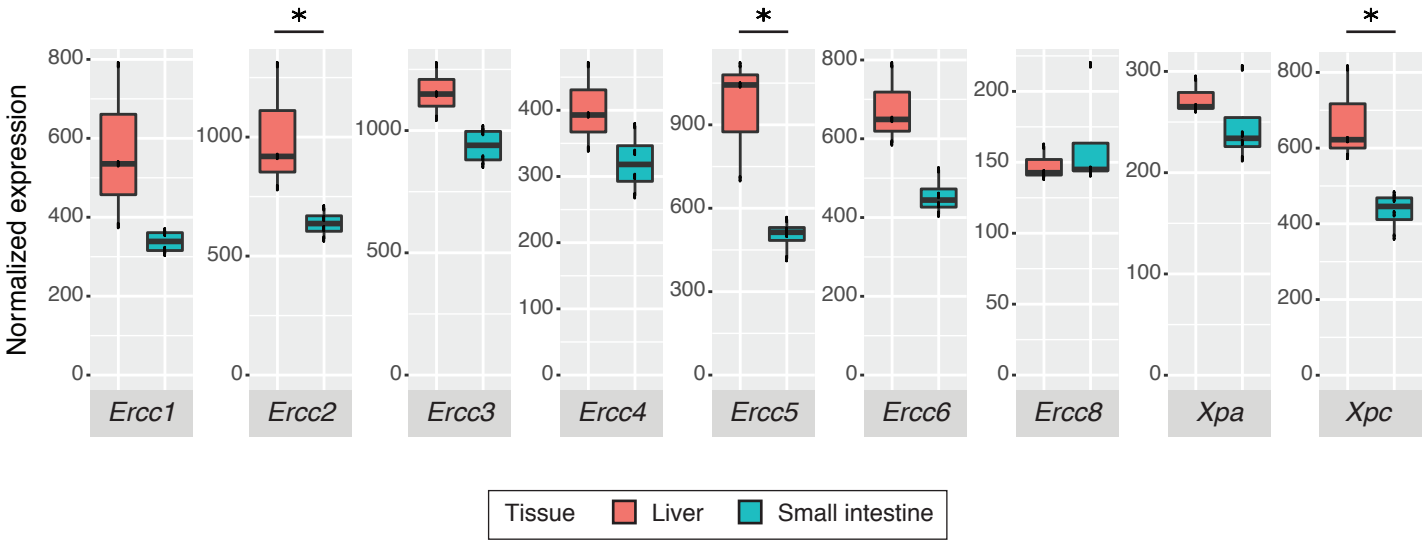
[7]These authors contributed equally to this work.

[*]Corresponding authors: R.vanBoxtel@prinsesmaximacentrum.nl, ecuppen@umcutrecht.nl

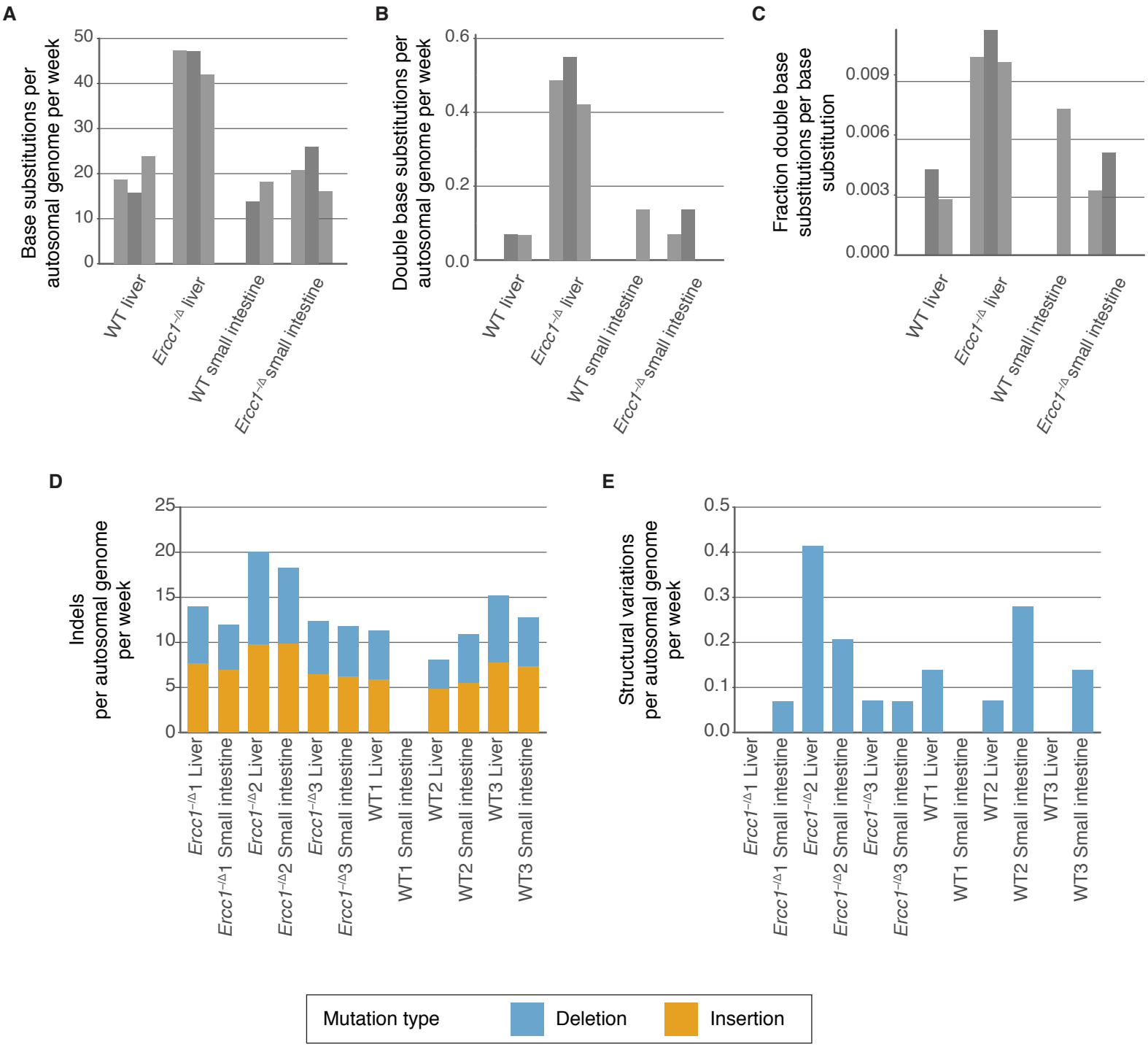
Table of Contents

Supplemental Figures:

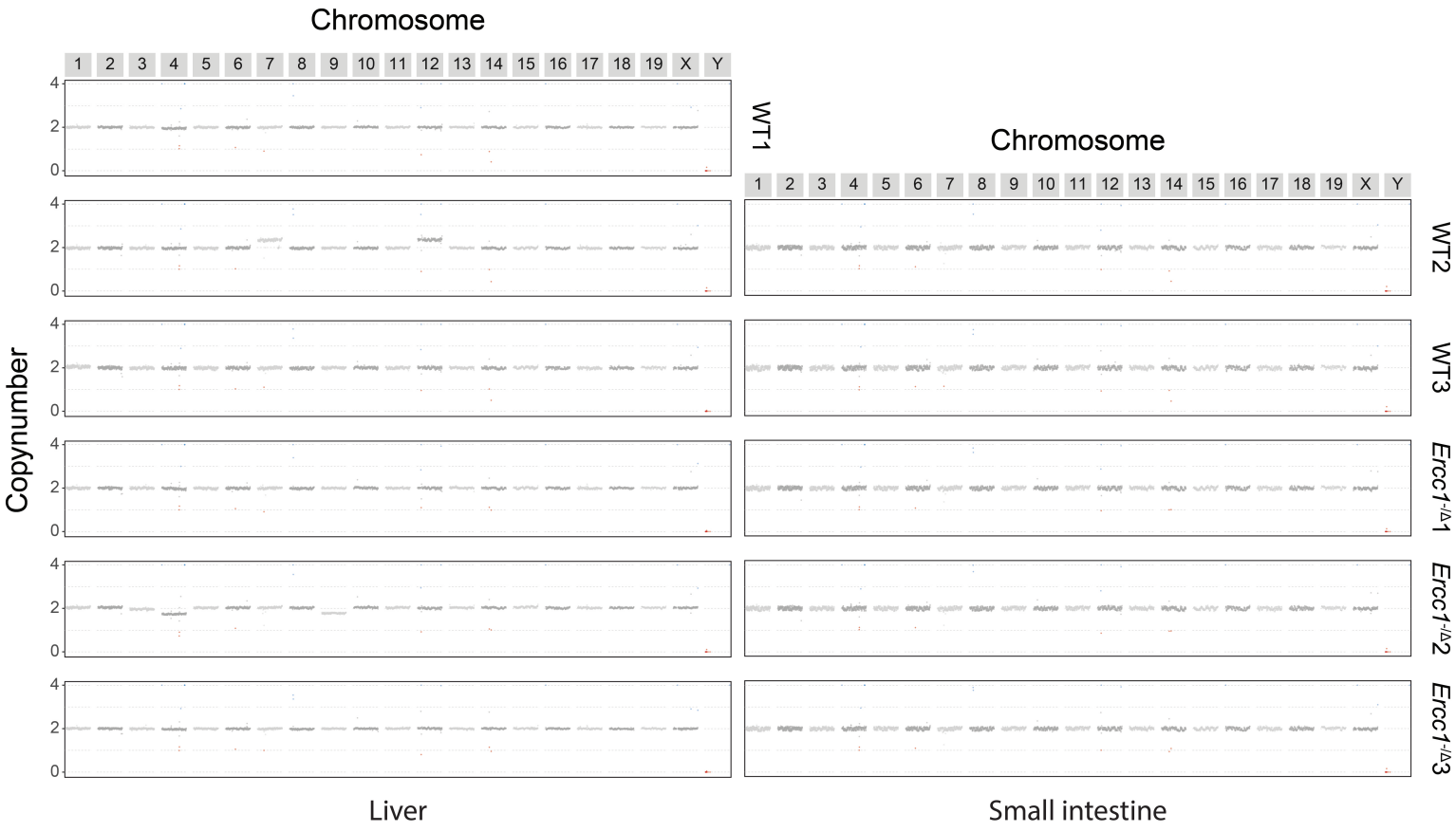
Supplemental Tables:



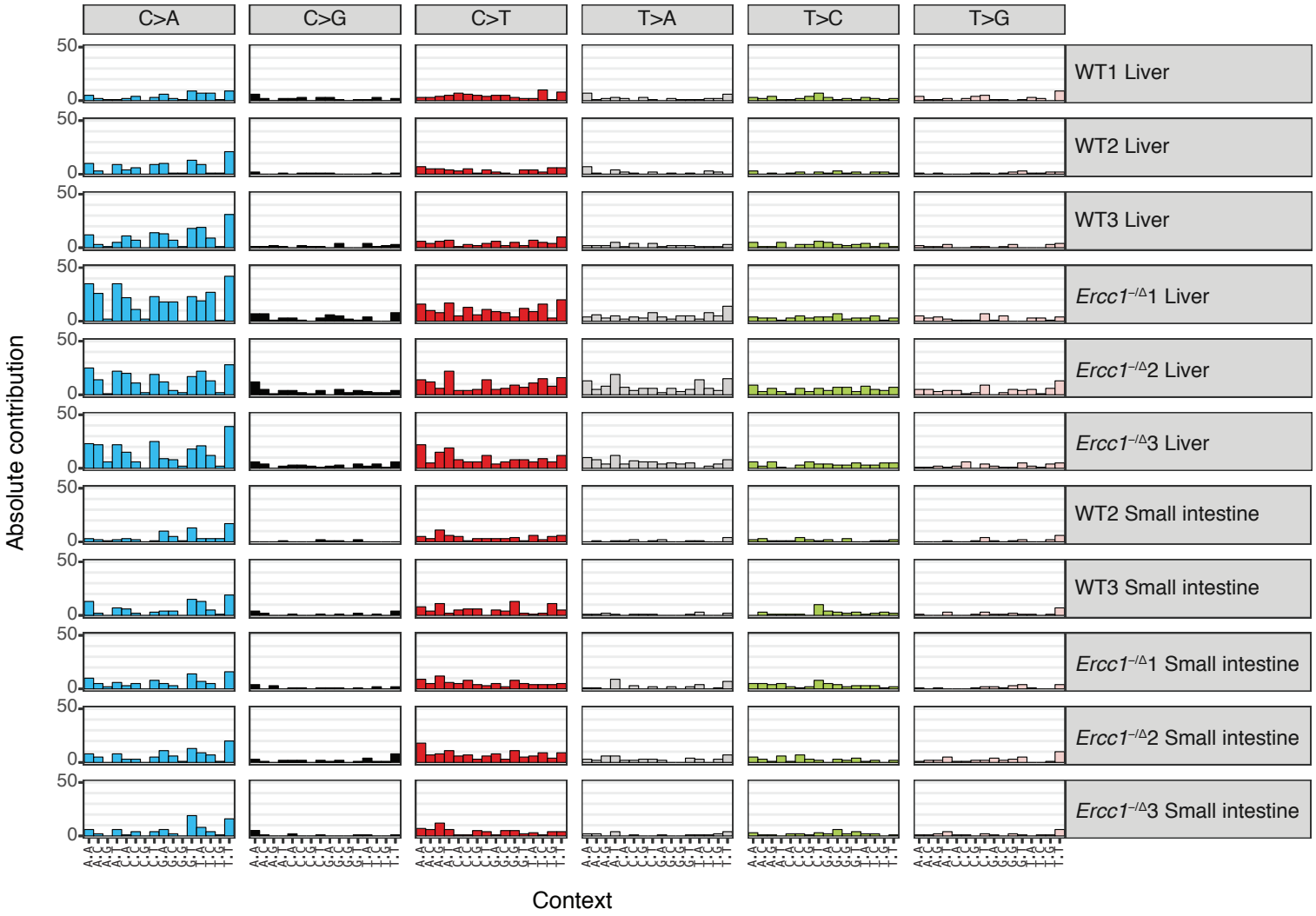
Supplemental Figure S1. Boxplots of normalized expression of 9 core NER genes in WT mouse ASCs from liver (n = 3) and small intestine (n = 4). Asterisks represent significant differential expression ($q < 0.05$, two-sided *t*-test, FDR correction).



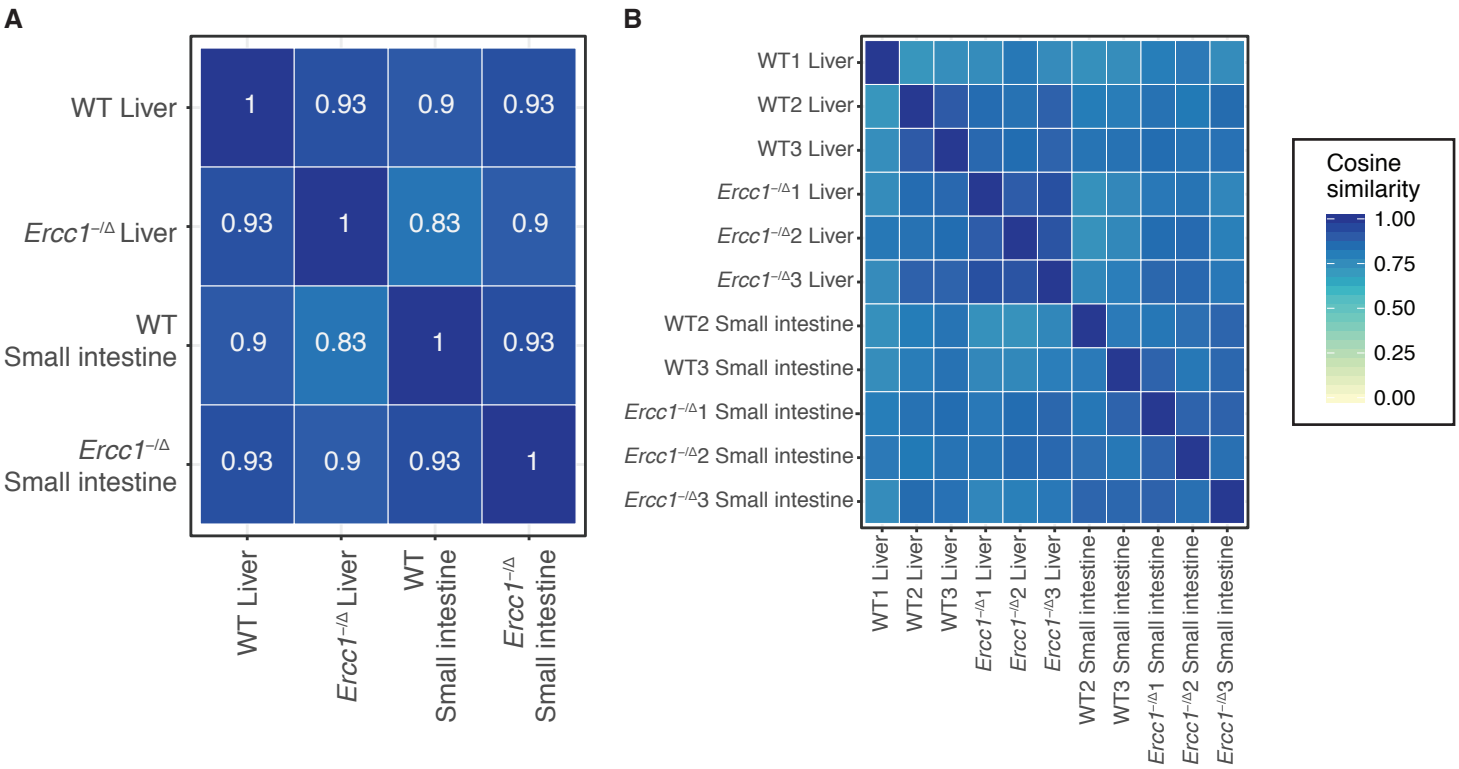
Supplemental Figure S2. Somatic mutation rates in the genomes of single ASCs from liver and small intestine of WT and *Ercc1*$^{-/\Delta}$ mice. (A) Base substitutions, (B) double base substitutions, (C) fraction double base substitutions/base substitutions, (D) indels, and (E) SVs acquired per autosomal genome per week in single ASCs from WT liver, *Ercc1*$^{-/\Delta}$ liver, WT small intestine, and *Ercc1*$^{-/\Delta}$ small intestine.



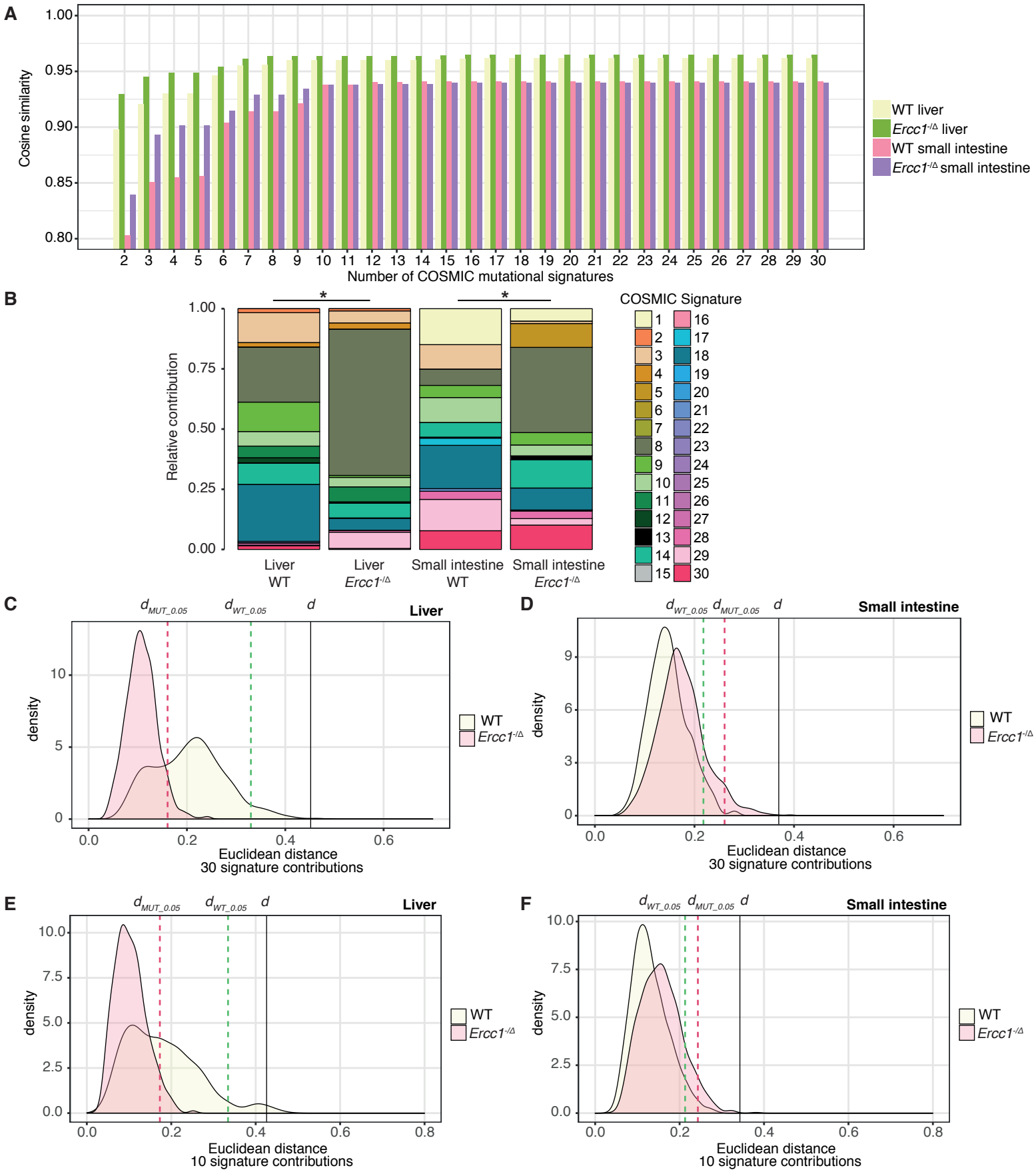
Supplemental Figure S3. Genome-wide copy-number profiles of single ASCs from liver and small intestine of WT and *Ercc1*$^{-/\Delta}$ mice.



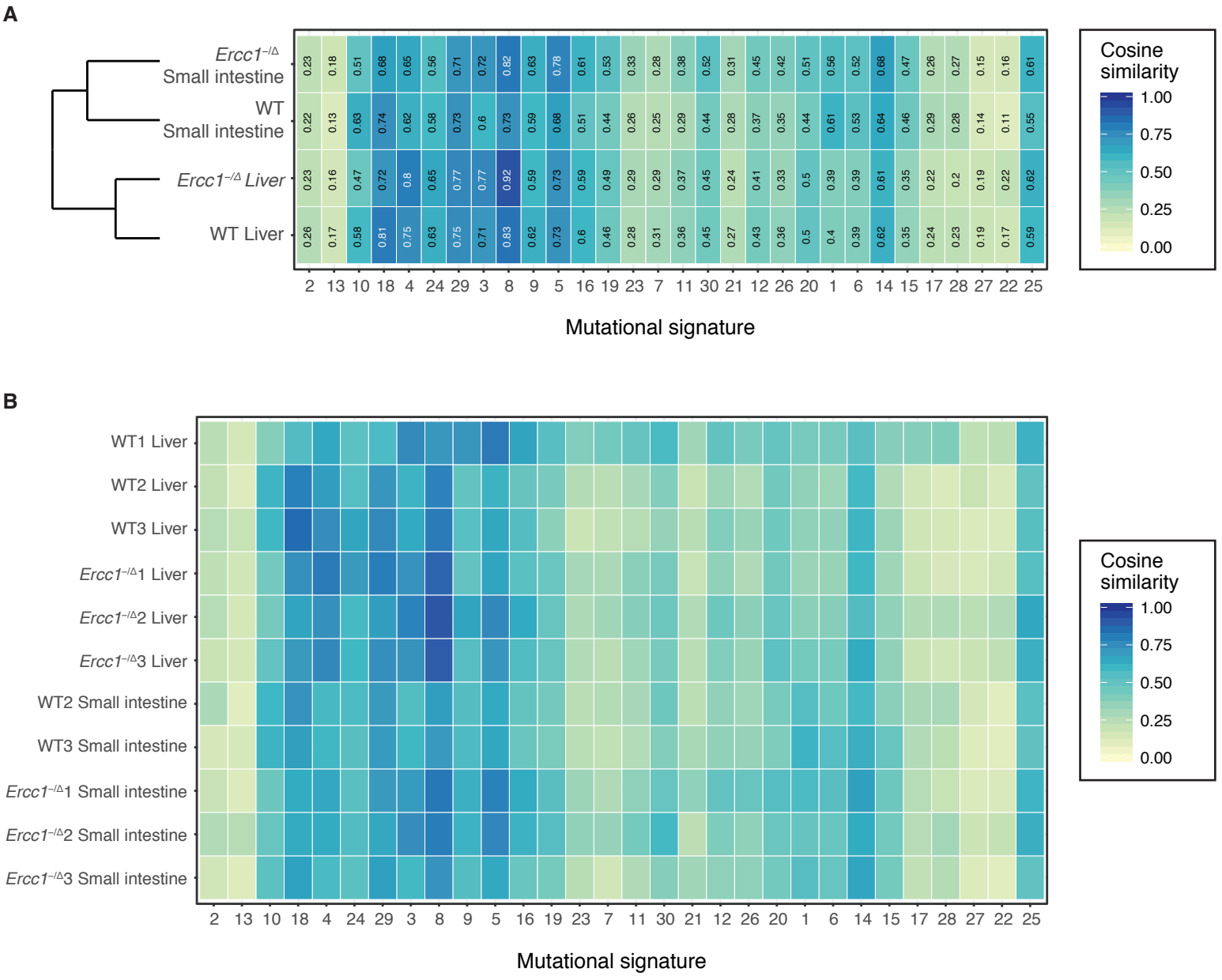
Supplemental Figure S4. Absolute contribution of each indicated context-dependent base substitution type to the mutational profiles of ASCs from liver and small intestine of WT and *Ercc1*$^{-/\Delta}$ mice.



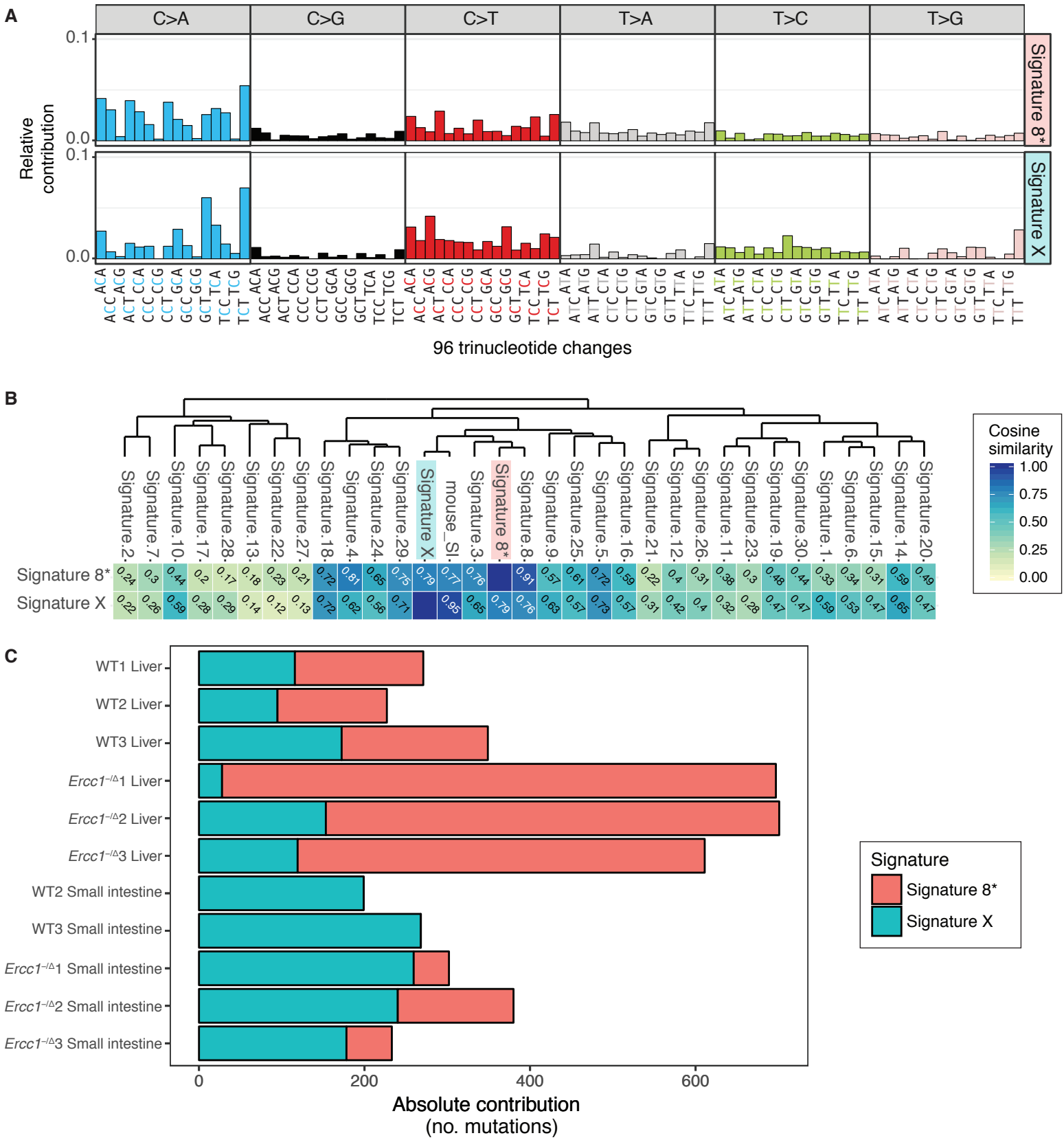
Supplemental Figure S5. Similarity between mutational profiles. (A) Cosine similarity between the mutational profiles of all indicated mouse ASC groups (B) Cosine similarity between the mutational profiles of all mouse ASCs.



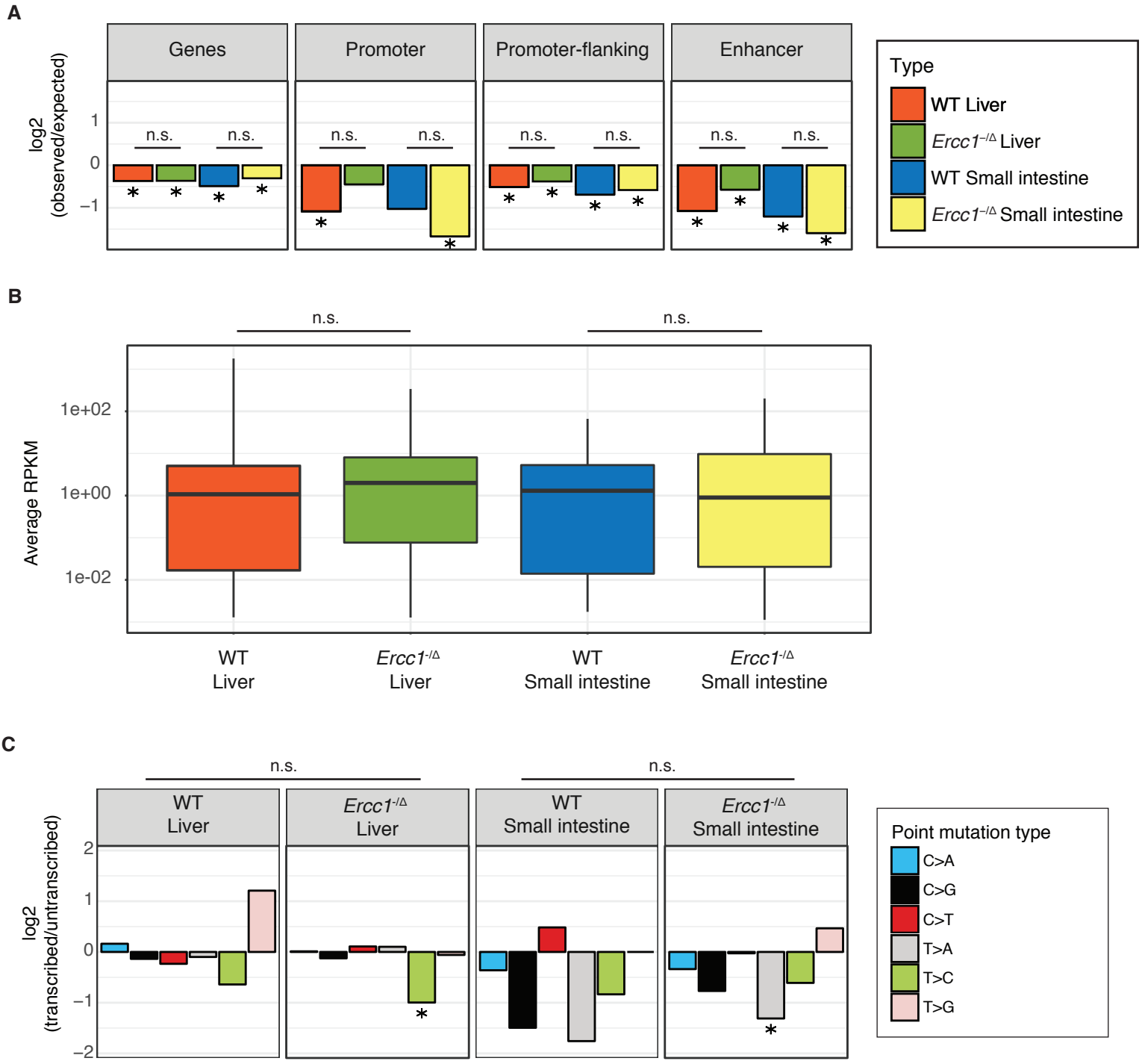
Supplemental Figure S6. (A) Cosine similarity between the original centroids and those that were reconstructed using the indicated number of COSMIC signatures for each ASC group. (B) Relative contribution of the 30 COSMIC mutational signatures to the centroids of each ASC group. Asterisks indicate significantly different signature contributions, as determined using the bootstrap distributions depicted in c and d. (C) A bootstrap resampling approach was taken to construct a population of WT and *Ercc1*$^{-/\Delta}$ samples. Subsequently, 3 replicas were randomly selected and the relative signature contributions were calculated for the centroid of the replicas. The Euclidean distance was calculated between this contribution vector and that of the original centroid. This was repeated 10,000 times to construct a distribution of distances d_{WT} for WT (shown in green). Similarly a distribution of d_{MUT} was generated for *Ercc1*$^{-/\Delta}$ (shown in red). The green dashed line indicates the distance where p value = 0.05, $d_{WT\_0.05}$, and red dashed line indicates the distance where p value = 0.05, $d_{MUT\_0.05}$. The distance between the relative signature contributions of the original WT and *Ercc1*$^{-/\Delta}$ centroids, d, is indicated with a black line. The signature contributions of WT and *Ercc1*$^{-/\Delta}$ are considered to be different when $d > d_{MUT\_0.05}$ and $d > d_{WT\_0.05}$. Similarly, bootstrap distributions were generated for small intestine (D), and for relative contributions of a subset of 10 COSMIC signatures for both liver (E) and small intestine (F).



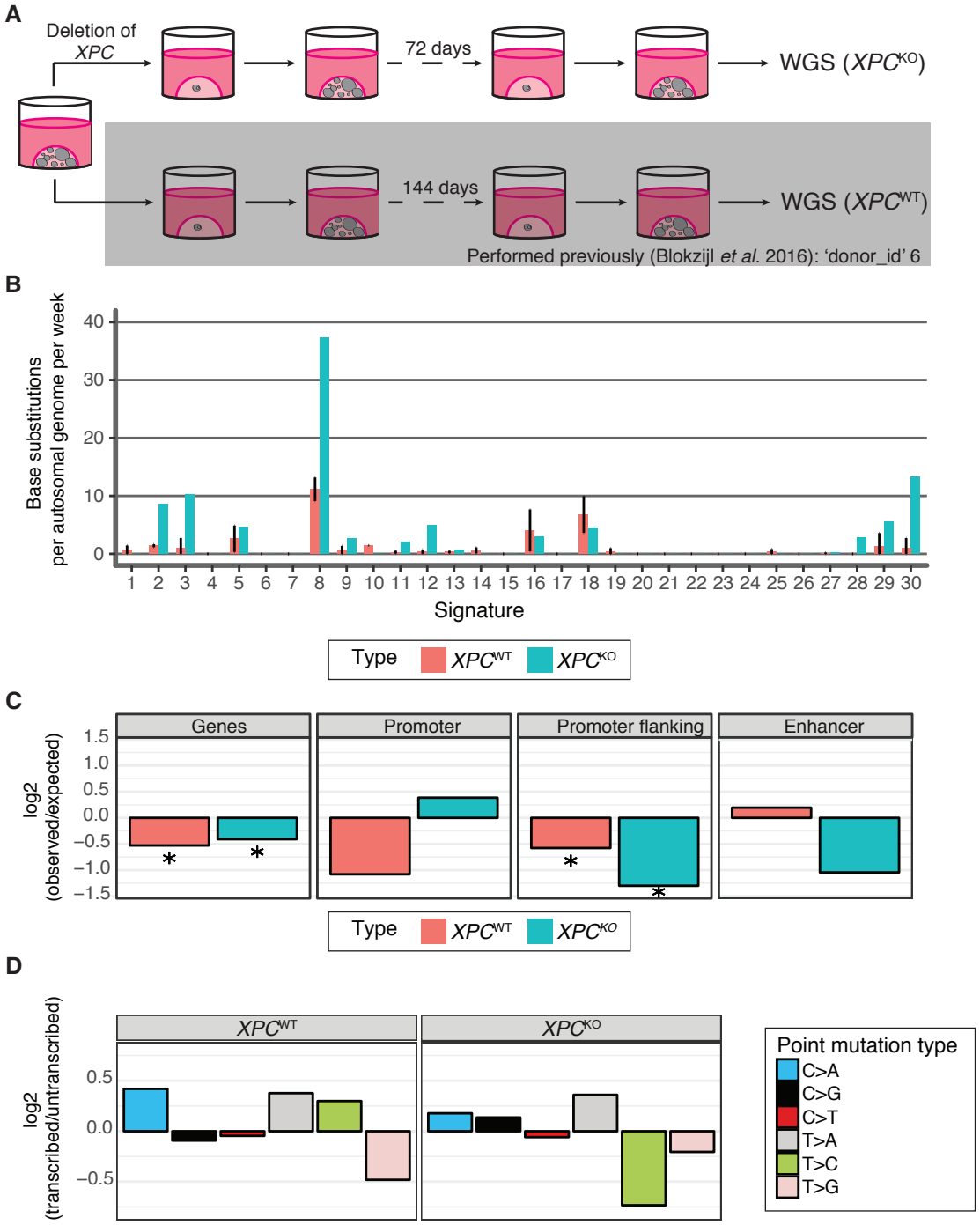
Supplemental Figure S7. Similarity between mutational profiles and mutational signatures. Cosine similarity between the mutational profiles of and each COSMIC mutational signatures. (A) per indicated mouse ASC group (B) per mouse ASC. The signatures have been ordered according to hierarchical clustering (complete linkage) using the cosine similarity between signatures, such that similar signatures are displayed close together. The samples are hierarchically clustered (complete linkage) in (A) using the Euclidean distance between the vectors of cosine similarities with the signatures.



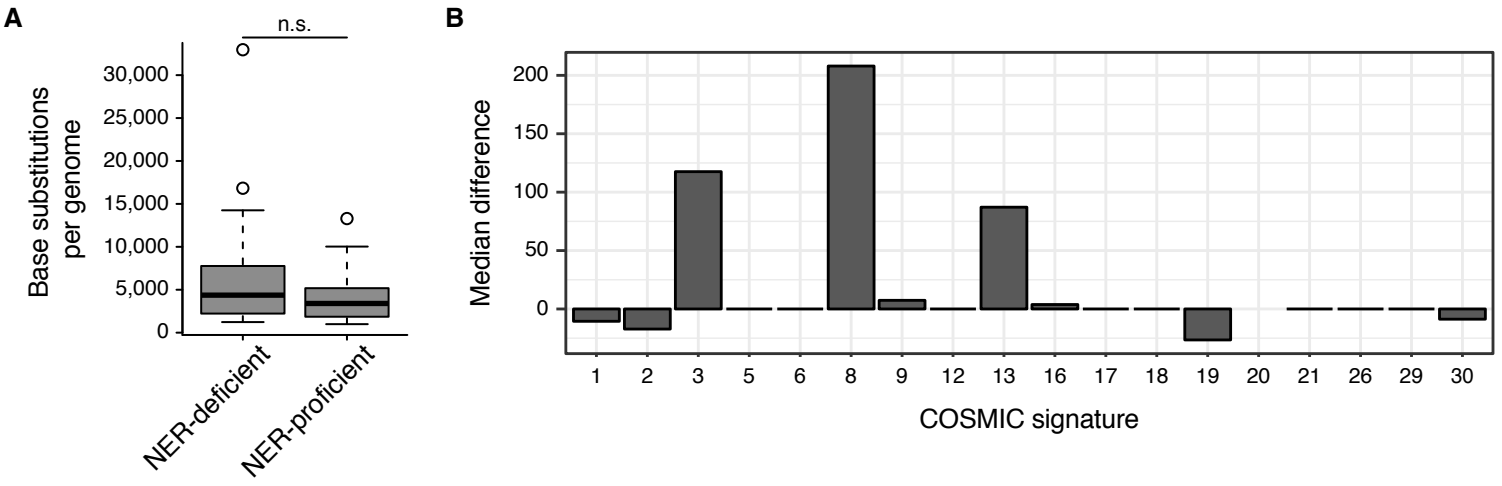
Supplemental Figure S8. Mutational signatures in mouse adult stem cells (ASCs). (A) Relative contribution of each indicated context-dependent base substitution type to the two mutational signatures that were extracted by non-negative matrix factorization (NMF) of the 96-channel mutational profiles of the mouse ASCs. (B) Cosine similarity of the two mutational signatures depicted in (A) to the 30 current COSMIC mutational signatures and to a centroid mutational profile detected in small intestinal ASCs of old mice (data published previously, Behjati *et al.* 2014). The signatures have been ordered according to hierarchical clustering (complete linkage) using the cosine similarity between signatures, such that similar signatures are displayed close together. (C) Absolute contribution of the two mutational signatures depicted in (A) to the mutational profiles of the sequenced mouse ASCs.



Supplemental Figure S9. Genomic distribution of somatic base substitutions the genomes of ASCs from WT liver (n = 3), *Ercc1*$^{-/\Delta}$ liver (n = 3), WT small intestine (n = 2), and *Ercc1*$^{-/\Delta}$ small intestine (n = 3). (A) Depletion of somatic base substitutions in genes, promoter, promoter-flanking regions, and enhancers for each indicated ASC group. The log2 ratio of the number of observed and expected base substitutions indicates the effect size of the depletion in each region. Asterisks represent significant depletions per indicated ASC group ($P < 0.05$, Binomial test, one-sided). n.s. : denotes non-significant differences in depletion between ASC groups ($q \geq 0.05$, Poisson test, two-sided). (B) Boxplots of the Reads Per Kilobase per Million mapped reads (RPKM) values of the genes in which a somatic SNV was detected per ASC group. n.s. : denotes non-significant differences in mean expression levels between ASC groups ($q \geq 0.05$, *t*-test, two-sided). (C) Transcriptional strand bias of base substitutions in genic regions. Log2 ratio of the number of mutations on the transcribed and untranscribed strand per indicated point mutation type for each sample. Asterisks represent significant strand asymmetries per indicated ASC group ($P < 0.05$, Poisson test, two-sided). n.s. : denotes non-significant differences in strand asymmetry between ASC groups ($q \geq 0.05$, Poisson test, two-sided).

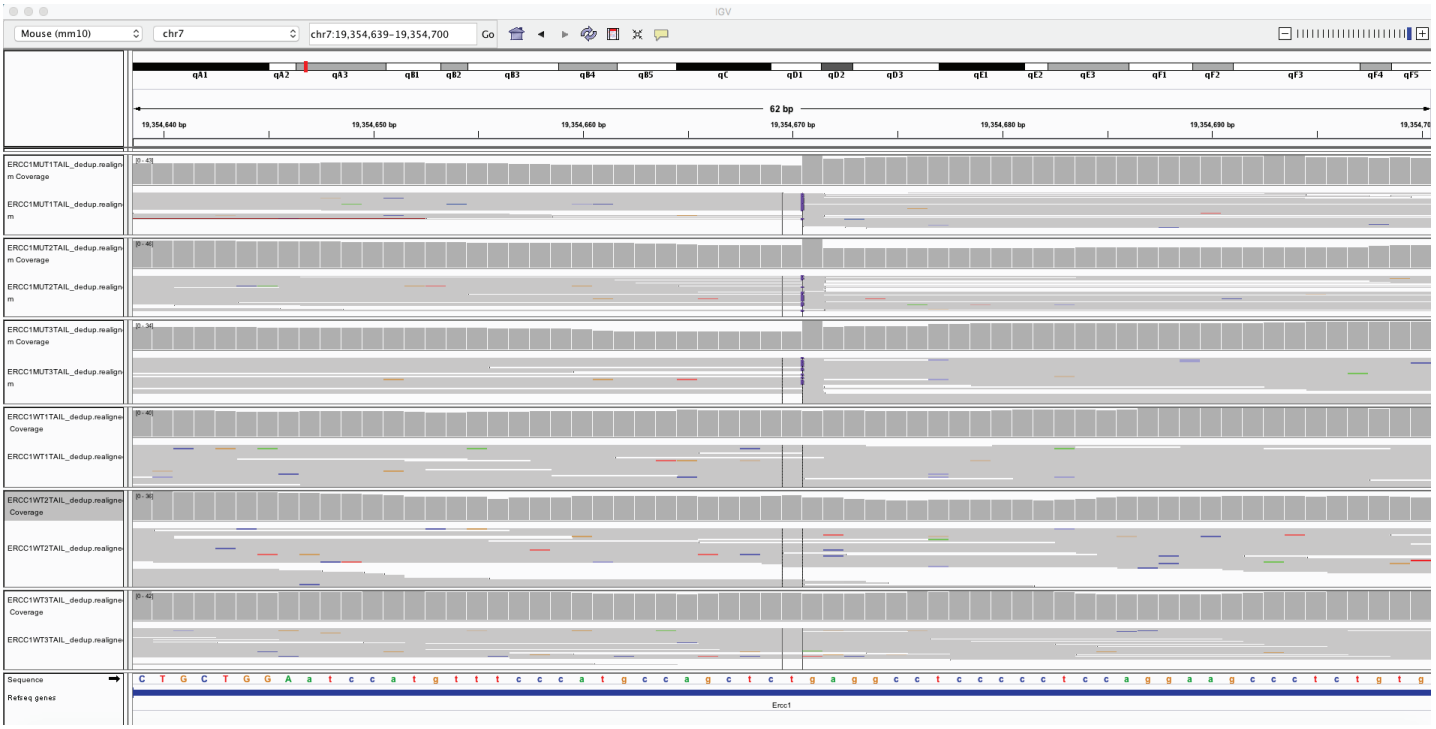


Supplemental Figure S10. Mutational consequences of deletion of *XPC* in human ASCs *in vitro*. (A) Schematic overview of the experimental setup used to determine the mutational consequences of KO of *XPC* in single ASCs. A clonal *XPC*^{KO} organoid culture was generated from a human organoid culture through CRISPR-Cas9 gene-editing. This organoid culture was cultured for 72 days to allow accumulation of sufficient mutations to perform downstream analyses. Subsequently, a subclonal organoid culture was derived from this clonal organoid culture and expanded until there was enough material to perform WGS. As a control sample for filtering germline variants, we used a blood sample that was genome sequenced previously (Blokzijl *et al.* 2016). The mutational patterns in the genome of *XPC*^{KO} ASCs were compared to mutational patterns observed previously in *XPC*^{WT} ASCs from the same human donor (Blokzijl *et al.* 2016a). (B) Contribution of the COSMIC mutational signatures to the mutational profile of *XPC*^{KO} and mean contribution of the COSMIC mutational signatures to the mutational profiles of *XPC*^{WT} ASCs. Error bars represent standard deviations. (C) Depletion/enrichment of base substitutions in genes, promoter, promoter-flanking regions, and enhancers. The log2 ratio of the number of observed and expected number of base substitutions indicates the effect size of the depletion/enrichment in each region. Asterisks represent significant depletions and enrichments per indicated ASC group ($P < 0.05$, Binomial test, one-sided). (D) Transcriptional strand bias of base substitutions in genic regions. Log2 ratio of the number of mutations on the transcribed and untranscribed strand per indicated point mutation type for each sample. Asterisks represent significant strand asymmetries per indicated ASC group ($P < 0.05$, Poisson test, two-sided).



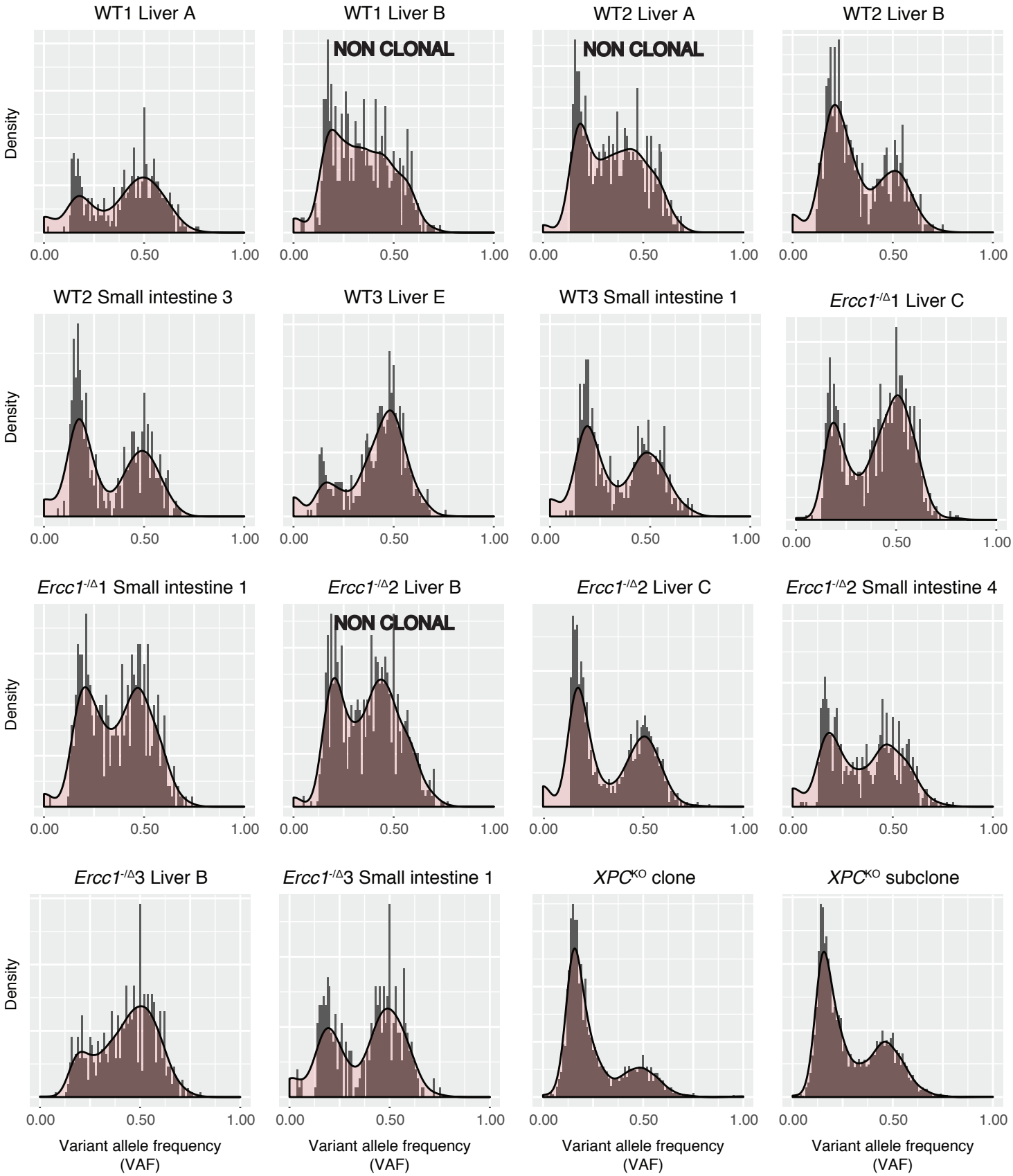
Supplemental Figure S11. (A) Boxplot depicting the total number of base substitutions in NER-deficient and NER-proficient breast cancer whole-genomes (n = 27 and n = 43, respectively). n.s. : non-significant ($P \geq 0.05$, Wilcoxon rank-sum test). (B) Median difference in the contribution of 18 COSMIC mutational signatures to the 96-channel mutational profiles of NER-proficient and NER-deficient breast cancer samples. Positive values indicate an increase in the contribution of a mutational signature in NER-deficient samples compared to NER-proficient samples.

A

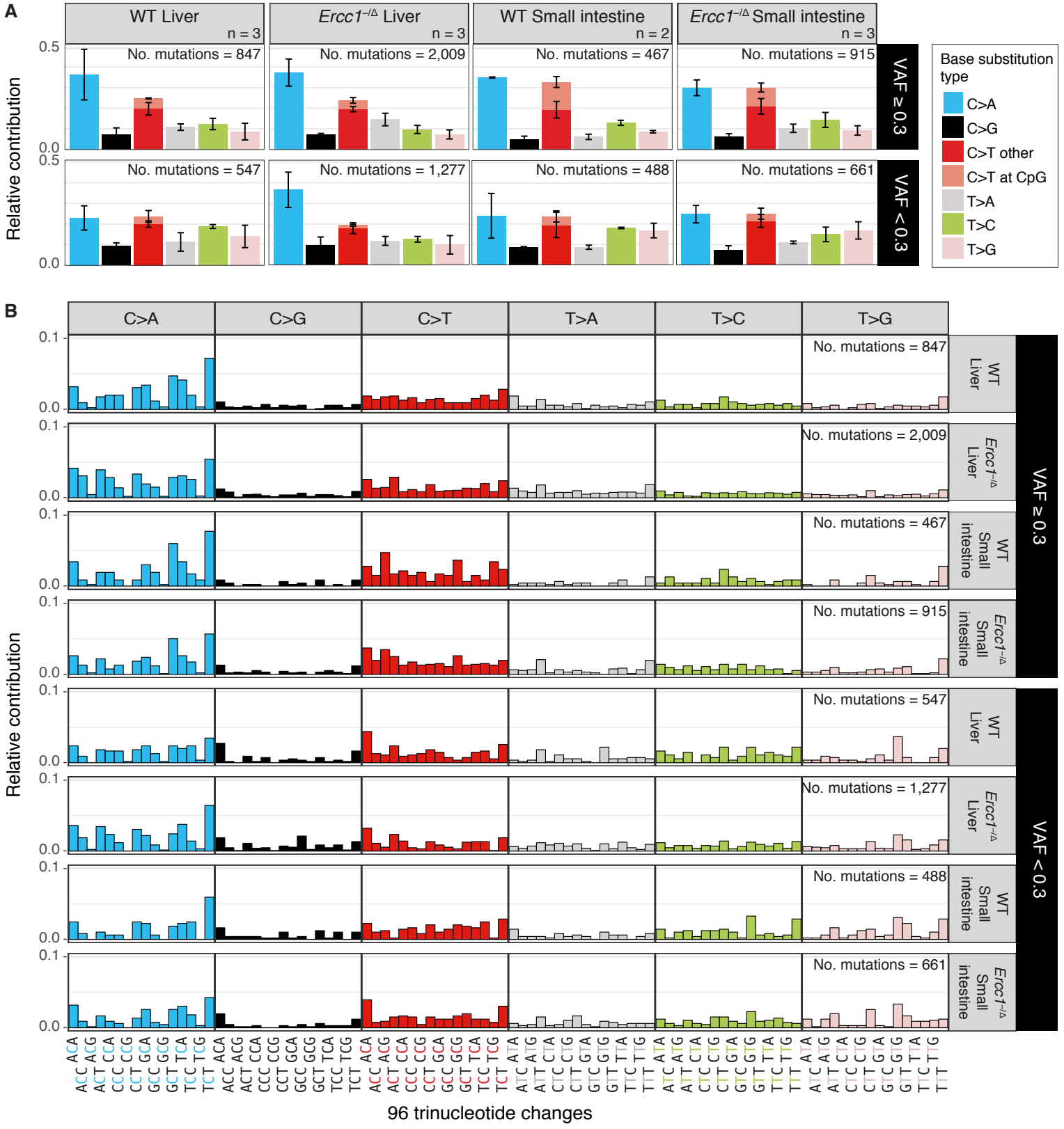


B

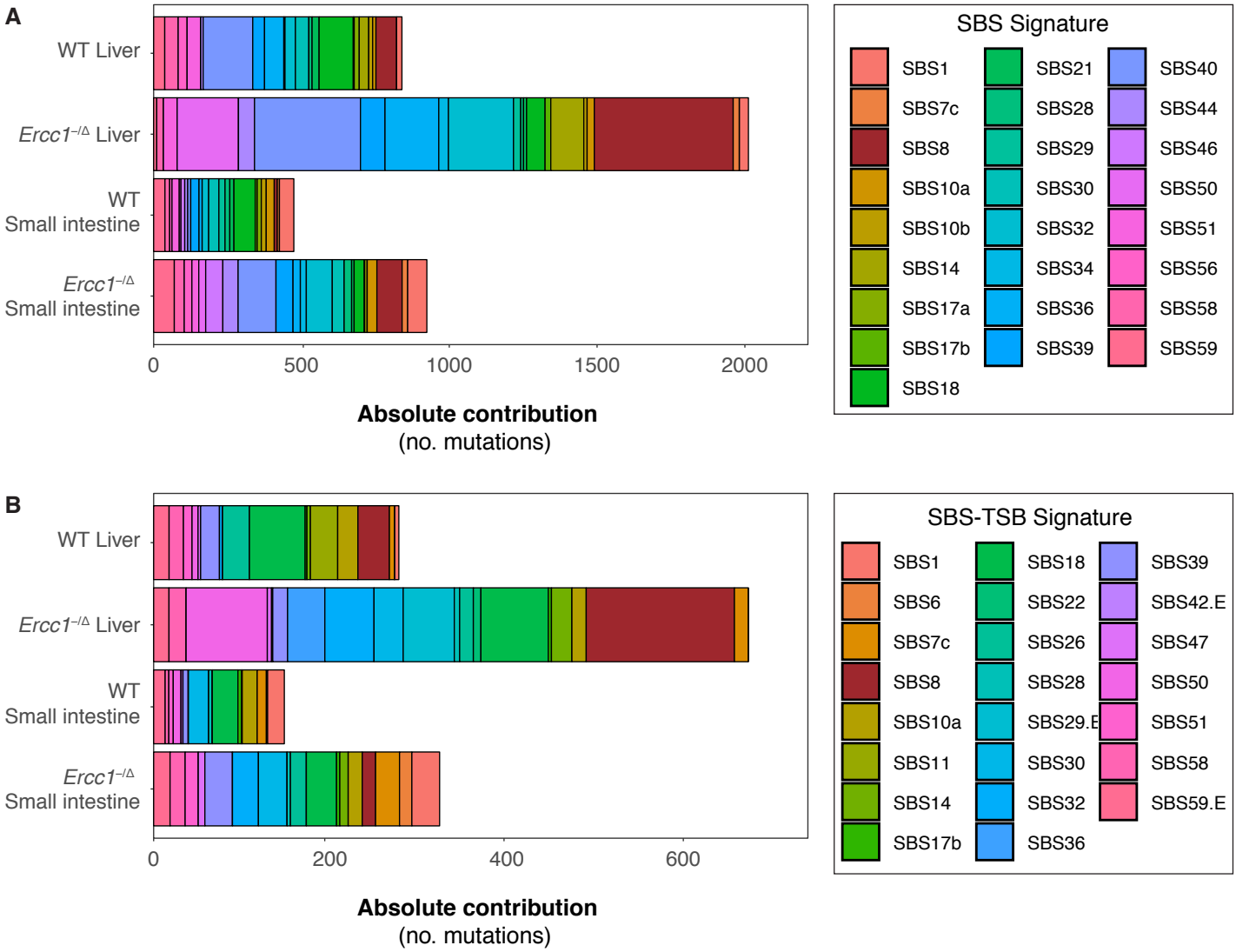
Supplemental Figure S12. IGV screenshots of mutations in the *Ercc1* gene in WT and *Ercc1*$^{-/\Delta}$ mice. (A) *Ercc1*$^{-}$ allele and (B) *Ercc1*$^{\Delta}$ allele in the WGS data of the tails of all WT and *Ercc1*$^{-/\Delta}$ mice.



Supplemental Figure S13. Distribution plot of the variant allele frequencies (VAFs) of all identified somatic base substitutions that remain before VAF ≥ 0.3 filtering for each ASC.



Supplemental Figure S14. Mean relative contribution of the indicated mutation types with a variant allele frequency (VAF) ≥ 0.3 and < 0.3 to the (A) 6-type mutation spectrum and (B) the 96-channel mutational profiles for each mouse ASC group. Error bars represent standard deviations. The total number of mutations and total number of ASCs (n) per group is indicated.



Supplemental Figure S15. Absolute contribution of the 60 SBS COSMIC mutational signatures (v3) to the base substitutions accumulated in the mouse ASCs. (A) Absolute contribution of the indicated SBS mutational signatures to the average 96-channel mutational profiles of each mouse ASC group. Signatures were obtained from https://www.synapse.org/#!Synapse:syn11726601/wiki/513478. (B) Absolute contribution of the indicated TSB mutational signatures to the average 96-channel mutational profiles within gene bodies of each mouse ASC group. TSB signatures were obtained from https://www.synapse.org/#!Synapse:syn11726601/wiki/513478.

Supplemental Table S1. The log2 fold-change in expression of 9 core NER genes between WT small intestinal ASCs and WT liver ASCs.

Ensembl gene ID	Gene symbol	log2FoldChange (WT SI/WT liver)	*P*-value	Adjusted *P*-value
ENSMUSG00000026048	*Ercc5*	-0.930	0.000	0.002
ENSMUSG00000030094	*Xpc*	-0.628	0.013	0.040
ENSMUSG00000030400	*Ercc2*	-0.656	0.009	0.040
ENSMUSG00000054051	*Ercc6*	-0.571	0.029	0.052
ENSMUSG00000003549	*Ercc1*	-0.747	0.027	0.052
ENSMUSG00000024382	*Ercc3*	-0.303	0.217	0.304
ENSMUSG00000022545	*Ercc4*	-0.321	0.236	0.304
ENSMUSG00000028329	*Xpa*	-0.151	0.529	0.595
ENSMUSG00000021694	*Ercc8*	0.138	0.601	0.601

SI = small intestine

Supplemental Table S2. Overview somatic base substitutions, indels, and structural variations detected in mouse ASCs.

Mouse	Tissue	Callable genome (%)	Base substitutions*	Double base substitutions*	Small insertions*	Small deletions*	Structural variants*
Ercc1 $^{-/\Delta}$1	Liver	90.0%	683	7	111	90	0
Ercc1 $^{-/\Delta}$1	Small intestine	90.2%	300	1	101	72	1
Ercc1 $^{-/\Delta}$2	Liver	90.8%	685	8	142	149	6
Ercc1 $^{-/\Delta}$2	Small intestine	90.7%	376	2	143	122	3
Ercc1 $^{-/\Delta}$3	Liver	89.0%	599	6	92	84	1
Ercc1 $^{-/\Delta}$3	Small intestine	90.0%	233	0	90	80	1
WT1	Liver	90.4%	271	0	86	78	2
WT1	Small intestine	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
WT2	Liver	89.1%	225	1	69	46	1
WT2	Small intestine	89.4%	199	0	79	77	4
WT3	Liver	90.7%	347	1	113	107	0
WT3	Small intestine	90.5%	264	2	107	78	2

* Observed number of mutations within the callable genome

Supplemental Table S3. Double base substitutions acquired in the genomes of WT and *Ercc1* $^{-/\Delta}$ mouse ASCs.

Mouse	Tissue	Chromosome	Position	Type
Ercc1 $^{-/\Delta}$1	Liver	1	19963547-19963548	CC>AT
Ercc1 $^{-/\Delta}$1	Liver	2	75869242-75869243	GG>AA
Ercc1 $^{-/\Delta}$1	Liver	2	97605553-97605554	GC>CT
Ercc1 $^{-/\Delta}$1	Liver	2	151610833-151610834	GG>TT
Ercc1 $^{-/\Delta}$1	Liver	10	49132495-49132496	TC>AA
Ercc1 $^{-/\Delta}$1	Liver	10	55373184-55373185	CT>TA
Ercc1 $^{-/\Delta}$1	Liver	17	3467736-3467737	GG>TT
Ercc1 $^{-/\Delta}$1	SI	17	83387913-83387914	GG>AA
Ercc1 $^{-/\Delta}$2	Liver	3	157520204-157520205	AA>TG
Ercc1 $^{-/\Delta}$2	Liver	5	59202560-59202561	GA>TT
Ercc1 $^{-/\Delta}$2	Liver	5	102337631-102337632	AG>GA
Ercc1 $^{-/\Delta}$2	Liver	6	111610726-111610727	AA>GG
Ercc1 $^{-/\Delta}$2	Liver	9	101601114-101601115	AC>GA
Ercc1 $^{-/\Delta}$2	Liver	10	40895371-40895372	TC>GT
Ercc1 $^{-/\Delta}$2	Liver	11	107811666-107811667	GC>TT
Ercc1 $^{-/\Delta}$2	Liver	14	50134277-50134278	TG>GT
Ercc1 $^{-/\Delta}$2	SI	4	18177691-18177692	AC>TT
Ercc1 $^{-/\Delta}$2	SI	13	54599043-54599044	CC>TT
Ercc1 $^{-/\Delta}$3	Liver	2	54244172-54244173	CC>AT
Ercc1 $^{-/\Delta}$3	Liver	4	39336415-39336416	CA>AC
Ercc1 $^{-/\Delta}$3	Liver	6	116403190-116403191	TC>GA
Ercc1 $^{-/\Delta}$3	Liver	11	70529426-70529427	TC>AA
Ercc1 $^{-/\Delta}$3	Liver	14	63208917-63208918	GC>AA
Ercc1 $^{-/\Delta}$3	Liver	15	14643203-14643204	TC>AA
WT2	Liver	10	107687236-107687237	AG>TT
WT3	Liver	1	13942075-13942076	AG>GT
WT3	SI	5	54402194-54402195	GT>TC
WT3	SI	14	13541219-13541220	CA>AT

Supplemental Table S4. SVs acquired in the genomes of WT and *Ercc1* $^{-/\Delta}$ mouse ASCs.

Mouse	Tissue	Chromosome	Start	End	Size (bp)	Type
Ercc1 $^{-/\Delta}$1	SI	14	98382845	98383374	529	deletion
Ercc1 $^{-/\Delta}$2	Liver	11	4307381	4308024	643	deletion
Ercc1 $^{-/\Delta}$2	Liver	11	96366839	96367238	399	deletion
Ercc1 $^{-/\Delta}$2	Liver	15	14954694	14961303	6609	deletion
Ercc1 $^{-/\Delta}$2	Liver	15	82986523	82989502	2979	deletion
Ercc1 $^{-/\Delta}$2	Liver	16	3744900	3745261	361	deletion
Ercc1 $^{-/\Delta}$2	Liver	19	25020360	25021085	725	deletion
Ercc1 $^{-/\Delta}$2	SI	3	108934215	108934569	354	deletion
Ercc1 $^{-/\Delta}$2	SI	4	88438548	88439859	1311	deletion
Ercc1 $^{-/\Delta}$2	SI	6	49000048	49000651	603	deletion
Ercc1 $^{-/\Delta}$3	Liver	15	98807785	98833375	25590	deletion
Ercc1 $^{-/\Delta}$3	SI	5	36712689	36713090	401	deletion
WT1	Liver	4	152179670	152523647	343977	deletion
WT1	Liver	17	52028043	52028582	539	deletion
WT2	Liver	19	14877486	14877950	464	deletion
WT2	SI	4	145065976	145066394	418	deletion
WT2	SI	5	41625866	41635804	9938	deletion
WT2	SI	5	41625866	41687721	61855	deletion
WT2	SI	17	35644289	35644686	397	deletion
WT3	SI	2	84747100	84747615	515	deletion
WT3	SI	6	139571588	139571908	320	deletion

bp = base pairs

Supplemental Table S5. Overview somatic base substitutions, indels, and structural variations detected in human ASCs.

Sample	No. weeks in culture	Callable genome (%)	Base substitutions*	Double base substitutions*
XPC^{WT}1	20.6	76.9%	467	3
XPC^{WT}2	20.6	75.9%	572	2
XPC^{WT}3	20.6	73.4%	503	2
XPC^{KO}	10.3	88.0%	895	23

* Observed number of mutations within the callable genome

Supplemental Table S6. Double base substitutions acquired in the genomes of *XPC* ^{WT} and *XPC* ^{KO} human ASCs.

Sample	Chromosome	Position	Type
XPC ^{WT}1	5	28200738-28200739	TG>CA
XPC ^{WT}1	14	86428191-86428192	TC>AA
XPC ^{WT}1	14	87517131-87517132	CC>AT
XPC ^{WT}2	1	218692585-218692586	CA>AT
XPC ^{WT}2	8	89594139-89594140	TC>GA
XPC ^{WT}3	3	147536195-147536196	AG>GA
XPC ^{WT}3	4	29048020-29048021	CC>AA
XPC ^{KO}	1	69460060-69460061	GA>AC
XPC ^{KO}	1	234938262-234938263	CA>TT
XPC ^{KO}	2	7379071-7379072	TA>GT
XPC ^{KO}	2	45682863-45682864	TT>AA
XPC ^{KO}	2	57337763-57337764	TT>AA
XPC ^{KO}	2	98545278-98545279	AC>GA
XPC ^{KO}	2	99366427-99366428	TG>AA
XPC ^{KO}	2	144580563-144580564	TG>CT
XPC ^{KO}	2	171495577-171495578	TC>GA
XPC ^{KO}	3	67720855-67720856	TA>AG
XPC ^{KO}	3	139045761-139045762	GG>AT
XPC ^{KO}	4	136997624-136997625	GG>AA
XPC ^{KO}	4	189948523-189948524	TC>AA
XPC ^{KO}	6	75466827-75466828	GA>TT
XPC ^{KO}	6	104498768-104498769	GT>AA
XPC ^{KO}	6	129297075-129297076	GT>AA
XPC ^{KO}	8	36646090-36646091	GT>AA
XPC ^{KO}	11	119899524-119899525	AC>TT
XPC ^{KO}	12	52842015-52842016	AC>GA
XPC ^{KO}	13	51476415-51476416	TT>GC
XPC ^{KO}	19	9559875-9559876	TC>GA
XPC ^{KO}	19	45595513-45595514	AC>TT
XPC ^{KO}	22	46424298-46424299	TC>AT

Supplemental Table S7. Culture media

	Mouse liver culture initiation medium	Mouse liver expansion medium	Mouse small intestine medium	Human small intestinal organoid medium
Adv+++ medium*	50%	90%	30%	30%
WNT3A conditioned medium**	35%	-	50%	50%
NOGGIN conditioned medium**	5%	-	10%	-
RSPOI conditioned medium**	5%	5%	10%	20%
B27	-	-	1x	-
B27 without retinoic acid	1x	1x	-	-
N2	1x	1x	-	-
Primocin	1x	1x	1x	1x
Nicotinamide	10 mM	10 mM	-	-
N-Acetylcysteïne	0.625 mM	0.625 mM	1.25 mM	1.25 mM
FGF-10	100 ng/ml	100 ng/ml	-	-
ROCKi	10 µM	-	10 µM	-
HGF	50 ng/ml	50 ng/ml	-	-
Gastrin	10 nM	10 nM	-	-
hEGF	50 ng/ml	50 ng/ml	50 ng/ml	50ng/ml
hES Cell Cloning & Recovery Supplement	-	-	1x	-
A83-01	-	-	-	0.5 µM
SB202190	-	-	-	10 µM
Recombinant Noggin	-	-	-	100 ng/ml

*Adv+++ medium: Advanced DMEM/F-12 + 1% GlutaMAX, HEPES 10 mM and 1% penicillin/streptomycin

** Produced in house