Supplementary Discussion

(A) Discussion of previous experiments that used somatic mutations of single cells to reconstruct embryological lineages.

Using somatic mutations to reconstruct the development of mice has previously been explored. In these experiments insertions and deletions in selected short tandem repeats were genotyped using PCR-based methods, typically from whole genome amplified DNA from single cells. In most, but not all, cases this experimental approach has been applied to cells derived from hypermutator mice in which mutational load at short tandem repeats is high.

Collectively, these studies have begun to demonstrate that reconstructing embryology from somatic mutations is feasible. Indeed, an advantage that this approach offers is that the same loci can be measured in different animals from the same species. A major disadvantage, however, is that false positive variants will be introduced into the data at several steps including whole genome amplification, PCR of the short tandem repeats and sequencing of the repeats. The extent of these errors, particularly from single cells, is not clear. Moreover, the potential intrinsic high mutability of short tandem repeats means that the genotypes at these loci cannot be assumed to be constant as cells evolve during embryogenesis. Thus mutations can revert and/or be obliterated by further mutations at the same repeat confusing phylogenetic analysis. Although these errors/issues in the data are usually statistically compensated for in data analysis, ultimately data precision is lost which precludes the exact insights into early embryogenesis that the approach we have taken, based on whole genome sequencing and substitutions, can provide. Of course, a further advantage of this approach is that it is based on unbiased sequencing for all classes of mutation and thus also informs on mutational processes that have been operative during embryogenesis.

(B) Discussion of the *in vivo* versus *in vitro* origin of mutations found in organoids.

We expanded single cells or monoclonal single glands / crypts *in vitro* to use the cellular machinery to amplify DNA. During the culturing period (3-6 weeks) it is likely that mutations have arisen *in vitro* (*in vitro* mutations). These mutations occur independently and privately in individual cells and present as subclonal when sequencing DNA derived from a culture. Therefore, when analysing mutations unique to each of the 25 organoids, we applied stringent filters to exclude subclonal variants and to target heterozygous variants. Heterozygous variants are present in every cell of the culture and therefore represent mutations of the stem cell that the culture was derived from. The stem cell had acquired these mutation *in vivo* (*in vivo* variants). Statistical analysis of the mutant

read frequency using the Dirichlet process confirmed that in every organoid presented, there is only a single clone which is heterozygous (Supplementary Figure 1). Thus, we are likely to have captured *in vivo* mutations in this experiment. This line of reasoning has an important limitation. Should a single cell, whilst in culture, gain a growth advantage over all other cells, it could in principle expand and become the major clone of the culture. With time, all cells in the culture would then be derived from this new dominant clone. Its mutations would comprise both *in vivo* mutations of the original stem cell as well as private mutations that this particular cell had acquired *in vitro*. Both sets of mutations would present as heterozygous.

Although it is conceivable that in our cultures aberrant clones with a growth advantage evolved, it would seem highly unlikely for this to have happened consistently and independently in every of the 25 organoid cultures during the limited culturing period. Further, one would expect that organoids dominated by a novel *in vitro* clone would differ in terms of mutation burden and spectrum. However, within each tissue in both mice the data is remarkably consistent in terms of mutations burden and spectrum. In small bowel we have been able to assess *in vitro* mutational spectrum directly by subcloning single organoid cells after eight weeks in culture. The spectrum of these *in vitro* mutations was markedly different from the spectrum of *in vivo* small bowel mutations (Supplementary Figure 7). Thus, overall the data indicate that the heterozygous mutations unique to each organoid represent *in vivo* mutations.

(C) Discussion of differences in mutation rate between embryonic and small bowel stem cells.

We estimated that small bowel stem cells accumulate ~1.1 mutations per cell division, similar to what we observe during embryogenesis (~1.5 mutations / division 35/23 cell divisions). Because these rates are very similar, and given the small number of embryonic divisions and mutations we are working with in this study, we cannot say whether there really is a difference between embryonic and adult stem cell mutation rates per cell division based on these calculations. It should also be noted that we cannot exclude additional early divisions due to limitations of sequencing coverage in the tail and thus statistical power. If there are more than 23 cell divisions in the two lineages, that would reduce the embryonic mutation rate from ~1.5 even further towards the rate of ~1.1 observed in adult small bowel stem cells.

Supplementary Table 1: Sample characteristics

ID	Alias ID	Age (weeks)	Site	Sequencing depth	Sensitivity*	Number of unique substitutions	Number of unique substitutions adjusted for sensitivity	C>T substitutions at XpCpG trinucleotides	C>T substitutions at XpCpG trinucleotides, adjusted for sensitivity
M1_Tail	MD4953b	116	Tail	37	1	-	-	-	-
M1_LB1	MD4953d	116	Colon	5	0.28	135	482	35	125
M1_LB2	MD4953d2	116	Colon	5	0.31	186	600	38	123
M1_LB3	MD4953d3	116	Colon	5	0.32	270	844	70	219
M1_LB4	MD4953d4	116	Colon	5	0.31	329	1061	58	187
M1_SB1	MD4953e	116	Small Intestine	8	0.48	452	942	63	131
M1_SB2	MD4953e2	116	Small Intestine	6	0.42	426	1014	92	219
M1_SB3	MD4953e3	116	Small Intestine	13	0.67	589	879	94	140
M1_SB4	MD4953e4	116	Small Intestine	5	0.36	246	683	46	128
M1_SB5	MD4953e5	116	Small Intestine	7	0.41	417	1017	73	178
M1_SB6	MD4953e6	116	Small Intestine	8	0.48	571	1190	79	165
M1_St1	MD4953f	116	Stomach	6	0.58	229	395	25	43
M1_St2	MD4953f2	116	Stomach	7	0.64	166	259	23	36
M1_St3	MD4953f3	116	Stomach	8	0.75	189	252	38	51
M2_Tail	MD4954b	98	Tail	35	1	-	-	-	-
M2_LB1	MD4954c	98	Colon	6	0.29	254	876	54	186
M2_LB2	MD4954c2	98	Colon	7	0.38	278	732	55	145
M2_LB3	MD4954c3	98	Colon	5	0.35	173	494	38	109
M2_SB1	MD4954d	98	Small Intestine	10	0.6	450	750	84	140
M2_SB2	MD4954d2	98	Small Intestine	6	0.38	326	858	52	137
M2_St1	MD4954e	98	Stomach	6	0.56	132	236	17	30
M2_St2	MD4954e2	98	Stomach	5	0.47	121	257	28	60
M2_St3	MD4954e3	98	Stomach	8	0.75	185	247	24	32
M2_P1	MD4954f	98	Prostate	5	0.44	117	266	3	7
M2_P2	MD4954f2	98	Prostate	6	0.58	168	290	13	22
M2_P3	MD4954f3	98	Prostate	6	0.53	95	179	13	25
M2_P4	MD4954f4	98	Prostate	6	0.5	210	420	16	32

Supplementary Table 2: Catalogue of embryonic mutations

WT: wildtype base; MT: mutant base (highlighted in apricot)

Mouse 1

Mutation ID (Fig 1A)	Embryonic cell ID (Fig 1A)	Cell generation (Fig 1A)	Chr	Position	Change	In pyrimidine context	TAIL variant depth	TAIL depth	TAIL variant allele fraction	Gene	Transcript	c.	p.	Consequence	LB1	LB2	LB3	LB4	SB1	SB2	SB3	SB4	SB5	SB6	ST1	ST2	ST3
1	**b**	**I**	9	119929275	C>T	C>T	12	32	0.375	Xirp1	ENSMUST0000011	-	-	Intronic	WT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT	MT
2	**c**	**I**	2	74353722	G>A	C>T	7	50	0.140	Intergenic	-	-	-	-	MT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
3	**c**	**I**	11	103413864	T>G	T>G	6	36	0.167	Gm884	ENSMUST0000005	-	-	Intronic	MT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
4	**d**	**II**	11	87609395	C>G	C>G	4	39	0.103	Mpo	CCDS25217.1	c.442C>G	p.R148G	Missense	WT	WT	MT	WT	WT	MT	WT	WT	MT	WT	WT	WT	WT
5	**d**	**II**	7	140000348	C>T	C>T	1	45	0.022	Fam53b	CCDS21926.1	-	-	Intronic	WT	WT	MT	WT	WT	MT	WT	WT	MT	WT	WT	WT	WT
6	**d**	**II**	5	52752445	C>T	C>T	0	28	0.000	Intergenic	-	-	-	-	WT	WT	MT	WT	WT	MT	WT	WT	MT	WT	WT	WT	WT
7	**d**	**II**	15	40224670	C>T	C>T	0	55	0.000	Intergenic	-	-	-	-	WT	WT	MT	WT	WT	MT	WT	WT	MT	WT	WT	WT	WT
8	**e**	**II**	9	6654521	G>T	C>A	20	43	0.465	Intergenic	-	-	-	-	WT	MT	WT	MT	MT	WT	MT	MT	WT	MT	MT	MT	MT
9	**e**	**II**	10	64673849	G>A	C>T	12	44	0.273	Intergenic	-	-	-	-	WT	MT	WT	MT	MT	WT	MT	MT	WT	MT	MT	MT	MT
10	**f**	**III**	3	138307960	T>A	T>A	0	15	0.000	Intergenic	-	-	-	-	WT	WT	WT	WT	WT	MT	WT	WT	MT	WT	WT	WT	WT
11	**f**	**III**	2	74493337	G>T	C>A	0	13	0.000	Evx2	CCDS16138.1	c.?	p.?	3'UTR variant	WT	WT	WT	WT	WT	MT	WT	WT	MT	WT	WT	WT	WT
12	**f**	**III**	2	101593078	G>T	C>A	0	23	0.000	Prr5l	CCDS38186.1	-	-	Intronic	WT	WT	WT	WT	WT	MT	WT	WT	MT	WT	WT	WT	WT
13	**h**	**III**	2	13170938	T>C	T>C	1	35	0.029	Rsu1	CCDS15693.1	-	-	Intronic	WT	MT	WT	MT	WT	WT	MT	WT	WT	MT	WT	WT	WT
14	**i**	**III**	3	136733227	C>T	C>T	13	44	0.295	Intergenic	-	-	-	-	WT	WT	WT	WT	MT	WT	WT	WT	MT	WT	WT	MT	MT
15	**i**	**III**	5	67365080	G>A	C>T	16	46	0.348	Limch1	CCDS39102.1	-	-	Intronic	WT	WT	WT	WT	MT	WT	WT	WT	MT	WT	WT	MT	MT
16	**i**	**III**	10	32999203	C>T	C>T	8	48	0.167	Intergenic	-	-	-	-	WT	WT	WT	WT	MT	WT	WT	MT	WT	WT	WT	MT	MT
17	**k**	**IV**	4	4477111	T>C	T>C	4	53	0.075	Intergenic	-	-	-	-	WT	WT	WT	MT	WT	WT	MT	WT	WT	MT	WT	WT	WT
18	**n**	**IV**	11	63339293	G>A	C>T	9	33	0.273	Intergenic	-	-	-	-	WT	WT	WT	WT	MT	WT	WT	MT	WT	WT	WT	WT	MT
19	**o**	**IV**	9	64102475	T>C	T>C	2	27	0.074	Intergenic	-	-	-	-	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	MT	MT	WT
20	**p**	**V**	11	3298140	C>A	C>A	0	30	0.000	Limk2	CCDS24360.1	-	-	Intronic	WT	WT	WT	MT	WT	WT	MT	WT	WT	WT	WT	WT	WT

Mouse 2

Mutation ID (Fig 1B)	Embryonic cell ID (Fig 1B)	Cell generation (Fig 1A)	Chr	Position	Change	In pyrimidine context	TAIL variant depth	TAIL depth	TAIL variant allele fraction	Gene	Transcript	c.	p.	Consequence	LB1	LB2	LB3	SB1	SB2	ST1	ST2	ST3	P1	P2	P3	P4
1	**b**	**I**	16	14653113	A>G	T>C	19	49	0.388	Intergenic	-	-	-	-	WT	MT	WT	MT	WT	MT	MT	WT	WT	MT	MT	MT
2	**b**	**I**	9	84768290	G>C	C>G	19	59	0.322	Intergenic	-	-	-	-	WT	MT	WT	MT	WT	MT	MT	WT	WT	MT	MT	MT
3	**b**	**I**	13	6137469	C>G	C>G	12	46	0.261	Intergenic	-	-	-	-	WT	MT	WT	MT	WT	MT	MT	WT	WT	MT	MT	MT
4	**c**	**I**	18	28624362	C>G	C>G	11	56	0.196	Intergenic	-	-	-	-	MT	WT	MT	WT	MT	WT	WT	MT	MT	WT	WT	WT
5	**c**	**I**	9	34199945	C>T	C>T	9	45	0.200	Intergenic	-	-	-	-	MT	WT	MT	WT	MT	WT	WT	MT	MT	WT	WT	WT
6	**c**	**I**	3	69509434	G>C	C>G	7	45	0.156	Intergenic	-	-	-	-	MT	WT	MT	WT	MT	WT	WT	MT	MT	WT	WT	WT
7	**c**	**I**	18	32603847	G>C	C>G	3	50	0.060	Intergenic	-	-	-	-	MT	WT	MT	WT	MT	WT	WT	MT	MT	WT	WT	WT
8	**c**	**I**	1	51529269	T>A	T>A	4	13	0.308	Obfc2a	CCDS35559.1	-	-	Intronic	MT	WT	MT	WT	MT	WT	WT	MT	MT	WT	WT	WT
9	**d**	**II**	1	79663252	A>G	T>C	4	19	0.211	Ap1s3	CCDS35628.1	-	-	Intronic	WT	MT	WT	MT	WT	MT	MT	WT	WT	WT	MT	MT
10	**e**	**II**	7	9589442	C>T	C>T	1	57	0.018	Intergenic	-	-	-	-	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	MT	WT	WT
11	**f**	**II**	9	123256166	G>A	C>T	5	23	0.217	Intergenic	-	-	-	-	MT	WT	WT	WT	WT	WT	WT	WT	WT	MT	WT	WT	WT
12	**i**	**III**	14	58588438	C>T	C>T	4	37	0.108	Efha1	CCDS27163.1	-	-	Intronic	WT	MT	WT	WT	WT	WT	WT	WT	WT	WT	MT	MT
13	**m**	**IV**	17	69045005	G>A	C>T	0	37	0.000	Intergenic	-	-	-	-	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	MT	MT
14	**m**	**IV**	1	184222456	G>A	C>T	0	29	0.000	Intergenic	-	-	-	-	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	MT	MT
15	**m**	**IV**	7	18647052	A>C	T>G	4	43	0.093	Intergenic	-	-	-	-	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	MT	MT