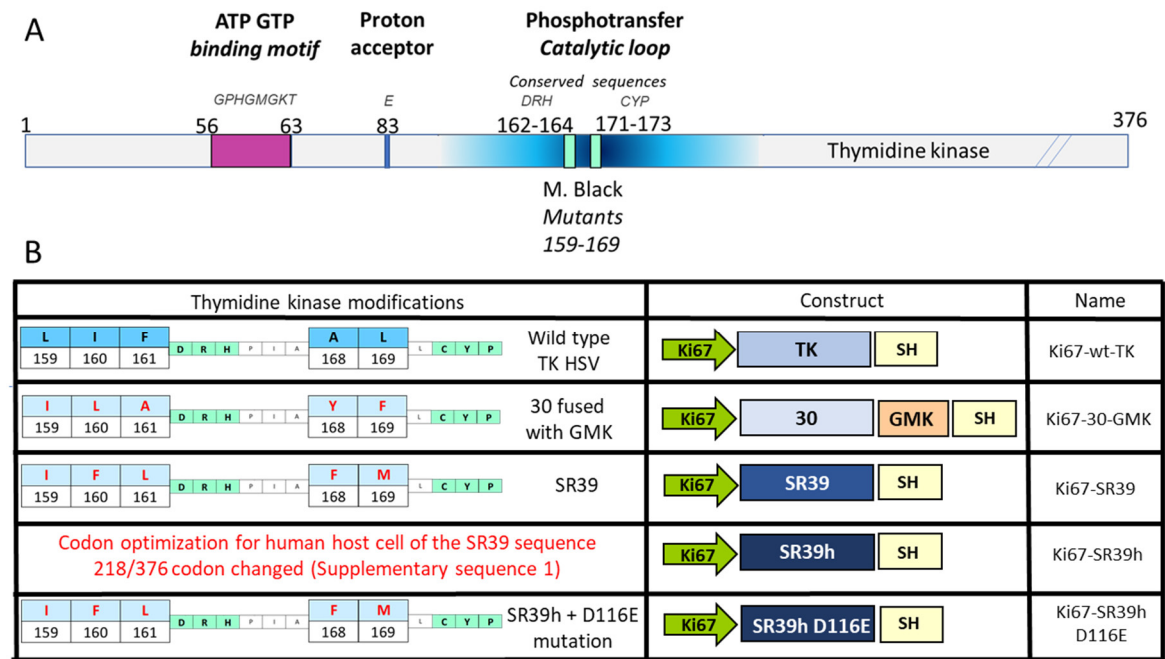
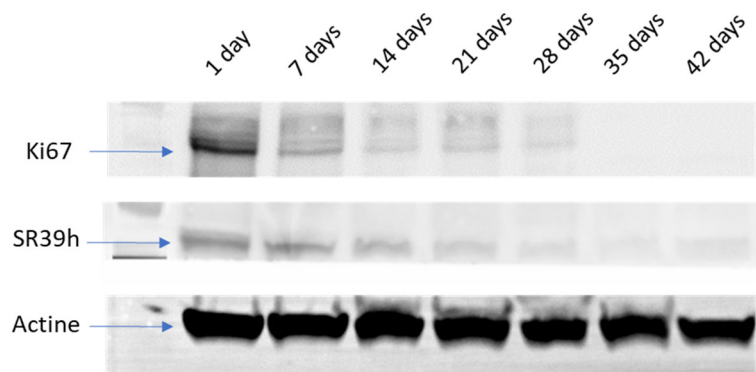


Supplementary data:

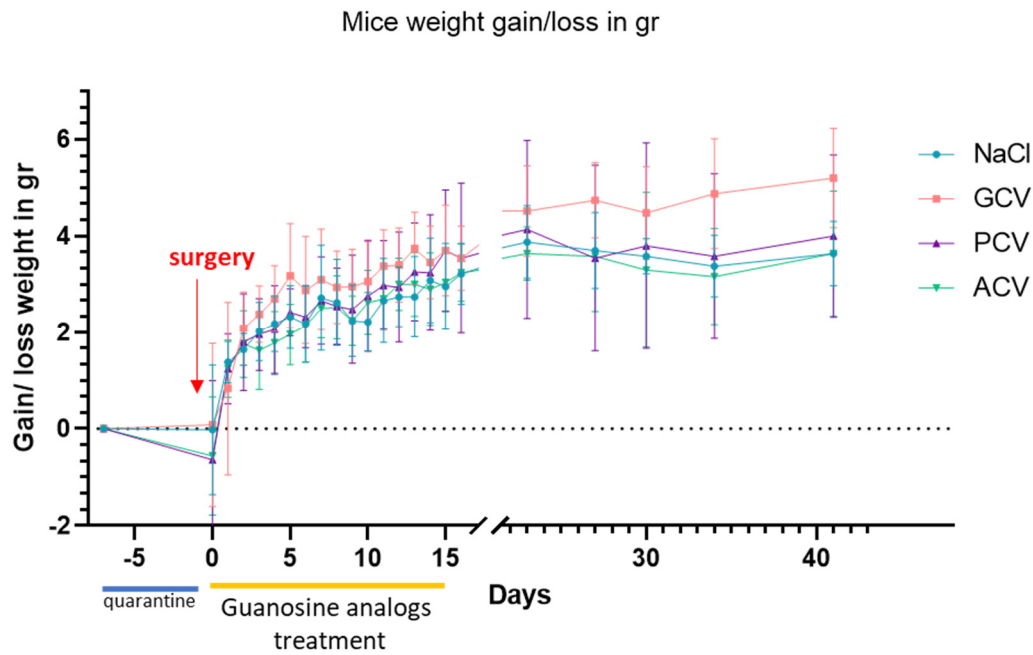


Supplementary Figure S1: Herpes simplex thymidine kinase (TK) variants used in this study.

A) Schematic representation of wild type TK; the two variants generated by Margaret Black (variant 30 and variant SR39) have amino changes close to the highly conserved region (162-164 and 171-173). B) Schematic representation of the different variants. Wild type TK is shown in the top lane, the two Margaret Black variants in lane 2 and 3, and two novel variants created for this study (SR39h and SR39h D116E) are shown in lane 4 and 5. Note that variant 30 was a fusion protein of a TK variant and guanylate kinase (GMK). The humanized SR39h variant codes for the same amino acid sequence as SR39. All variants had an N-terminal “she ble” protein (SH) extension conferring zeocin resistance to cell lines expressing the construct. For most experiments shown in our study, TK variants were expressed under the control of the Ki67 cell-cycle-dependent promoter.

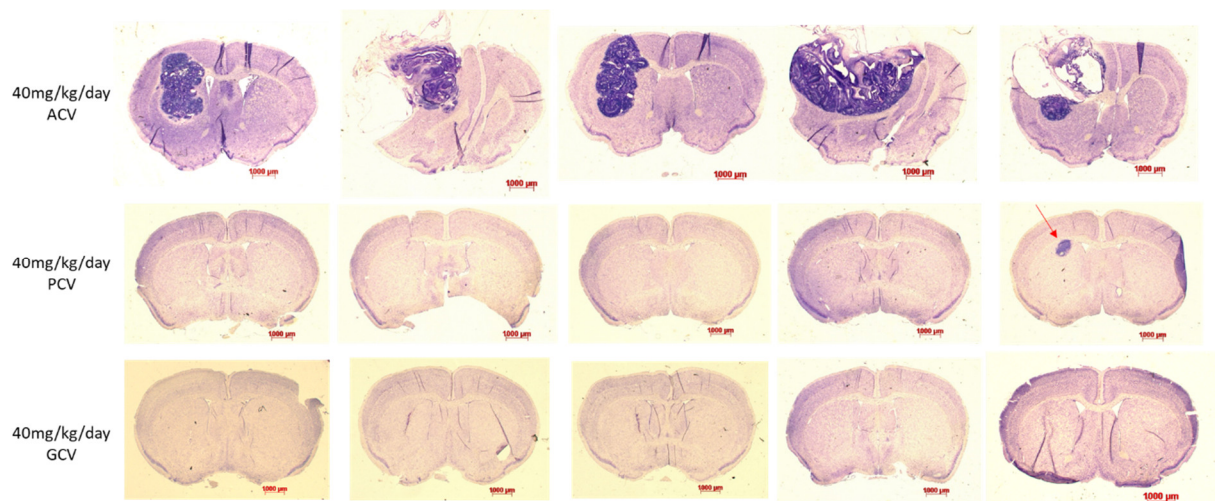


Supplementary Figure S2: Decreased of SR39h and Ki67 protein level during cell differentiation toward neurons. hESC expressing SR39h under the control of the Ki67 promoter were differentiated toward neurons during 42 days. Protein expression of SR39h and the endogenous Ki67 was analyzed by Western blot. The lower panel is a loading control with an anti-actin antibody.

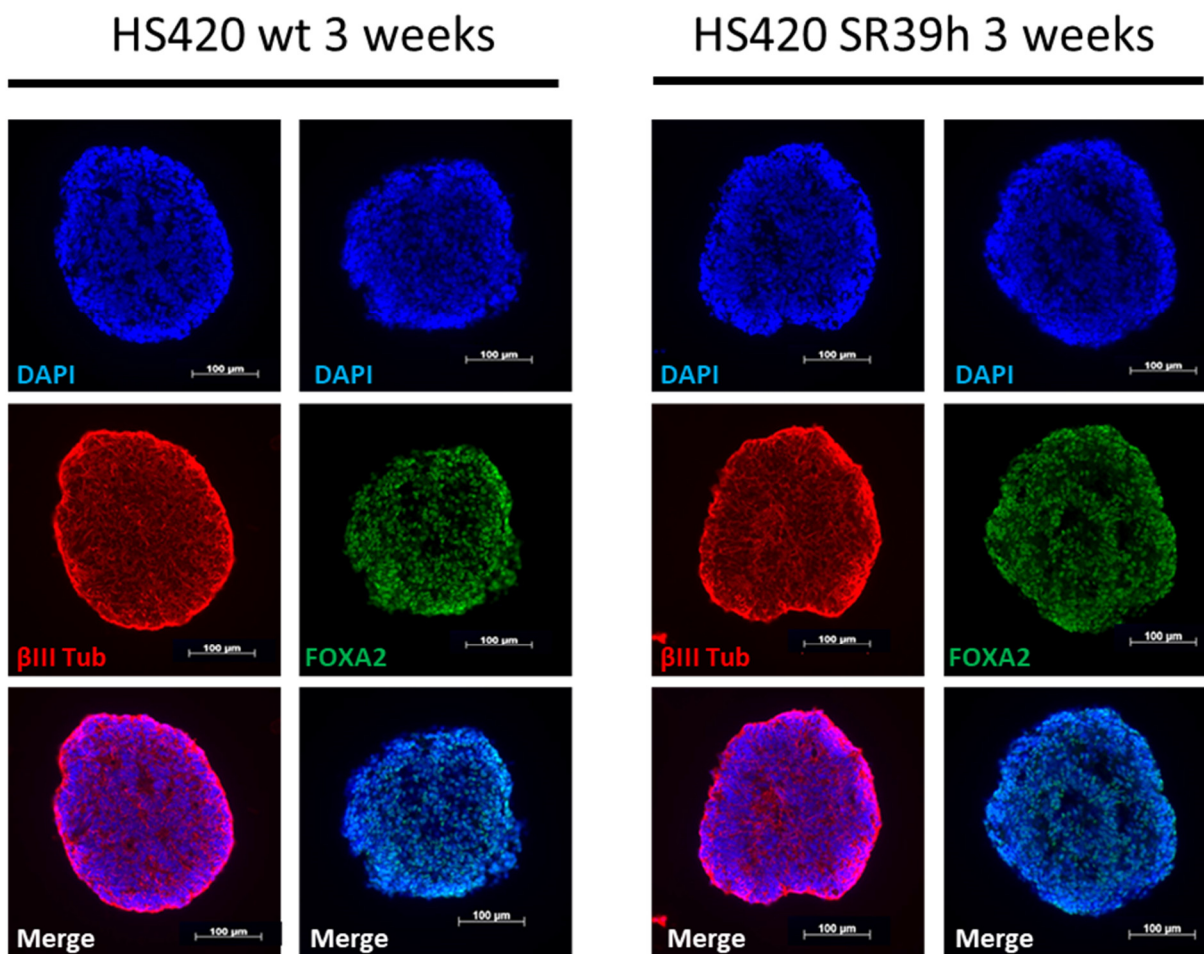


### Supplementary Figure S3: Mice weight

Mice were weighed daily during guanosine analogs treatment to identify differences in weight gain/loss between treated groups. No differences were observed between treatment groups.



**Supplementary Figure S4: *in vivo* validation of suicide gene.** Coronal section of NOD/SCID mice 47 days after injection of hESC expressing the SR39h thymidine kinase variant. Mice were treated for 2x5 days with the respective nucleoside analogs (see Methods). Upper lane: ACV; middle lane: PCV; lower lane: GCV. Sections were colored with cresyl violet and dark purple staining reflects a high density of nuclei. Scale bar 1000  $\mu$ m.



**Supplementary Figure S5: *hESC* differentiation toward neurons.**

Wild type and SR39h expressing hESC were differentiated towards neurons during 3 weeks. Immunostaining: FOXA2 (mouse, Santa Cruz, sc-374376), beta-III tubulin (βIII Tub, mouse, Sigma, T8660), nuclei were counterstained with DAPI.

**Supplementary sequence 1: SR39h and SR39 sequence alignment.** Viral-TK DNA sequence was humanized using web-based algorithms ([www.genscript.com](http://www.genscript.com)), nucleotides changed in SR39h sequence were labeled in red.

```
>>SR39 1131 bp (1131 nt)
Waterman-Eggert score: 3359; 221.0 bits; E(1) < 3.7e-61
77.5% identity (77.5% similar) in 1129 nt overlap (1-1129:1-1129)
```

	10	20	30	40	50	60
SR39h	ATGGCCTCTTATCC	TGGACACAGCACGC	CAGCGCCTTGATC	CAGGCTGCCAGATCT	AGAG	
	.....	.....	.....	.....	.....	.....
SR39	ATGGCTTCGTACCC	CGGCCATCAACACG	CGTCTGCGTTCGAC	CAGGCTGCGCGTT	CTCGC	
	10	20	30	40	50	60

	70	80	90	100	110	120
SR39h	GGCCACAGCAACAG	AGAGAACAGCCCT	GCCTCGGAGACAG	CAAGAGGCTACAG	AAGTT	
	.....	.....	.....	.....	.....	.....
SR39	GGCCATAGCAACCG	ACGTACGGCGTTG	CGCCCTCGCCGGC	AGCAAGAGCCACG	GGAAGTC	
	70	80	90	100	110	120

	130	140	150	160	170	180
SR39h	CGGCCC	GAGCAGAA	GATGCC	CACACT	GCTGAG	AGTGTACAT
SR39	CGCCCG	GAGCAGAAA	ATGCCAC	GCTACT	GCGGGTTT	TATATAGAC
	130	140	150	160	170	180
	190	200	210	220	230	240
SR39h	GGCAA	GACCAC	AACAACA	GCTGCT	GGTGGCC	CTGGGCAG
SR39	GGGAAA	ACCACC	ACGCAAC	TGCTGGT	TGGCCCT	TGGGTT
	190	200	210	220	230	240
	250	260	270	280	290	300
SR39h	GTGCCC	GAGCCT	TATGAC	CTATT	TGGAG	AGTGTG
SR39	GTACCC	GAGCCG	ATGACT	TACTGG	CGGGT	GCTGGG
	250	260	270	280	290	300
	310	320	330	340	350	360
SR39h	TACACC	CACACA	GCACCG	GCTGGAT	CAGGGC	GAATTT
SR39	TACACC	CACACA	ACACCG	CCTCGAC	CAGGGT	GAGATAT
	310	320	330	340	350	360
	370	380	390	400	410	420
SR39h	ATGACA	TCTG	CCAGAT	CACCAT	TGGGCAT	GCCTTAC
SR39	ATGACA	AAGCGC	CCAGATA	AACAAT	TGGGCAT	GCCTTAT
	370	380	390	400	410	420
	430	440	450	460	470	480
SR39h	CCTCA	CATTGG	CGGAGA	AGCGG	ATCTT	CTCATG
SR39	CCTCA	TATCGG	GGGGG	GAGGCT	GGGAGC	TACATG
	430	440	450	460	470	480
	490	500	510	520	530	540
SR39h	CTGGAC	AGACAC	CCCTAT	ATCGC	CTTCAT	GCTGTGT
SR39	CTCGAC	CGCCAT	CCCATC	GCCTT	CATGCT	GTGCTAC
	490	500	510	520	530	540
	550	560	570	580	590	600
SR39h	AGCATG	ACCTC	AGGCCG	TGCTGG	CTTTTC	GTGGCC
SR39	AGCATG	ACCCCC	AGGCCG	TGCTGG	CGTTTC	GTGGCC
	550	560	570	580	590	600
	610	620	630	640	650	660
SR39h	ACCAAT	TATCGT	GTGGG	AGCCCT	GCCTG	AGGACC
SR39	ACCAAC	ATCGTG	CTTGGG	GGCCCT	TCCGGAG	GACAGAC
	610	620	630	640	650	660
	670	680	690	700	710	720
SR39h	CAGCG	GCCTGG	CGAGAG	ACTGG	ATCTGG	CTATGCT

[illegible]