

Additional data file 2

Figure s1. *Rab3A* amplification in cDNA and genomic DNA (gDNA) samples. (a) Schematic representation of *Rab3A* gene and two splicing forms referred to as the long transcript (LT) and the short transcript (ST). The positions of primers used for RT-PCR are represented as arrows. (b) Comparison of PCR amplicons in cDNA and gDNA samples. Lane 1: cDNA; F1/R2. Lane 2: cDNA; F2/R2. Lane 3: gDNA; F1/R2. Lane 4: gDNA; F2/R2. Lane 5: cDNA; F1/R1. Lane 6: cDNA; F2/R1. Lane 7: gDNA; F1/R1. Lane 8: gDNA; F2/R1.

Figure s2. Screen of functional elements in the non-coding region of *Rab3A* locus. (a) Gene structure of *Rab3A* is illustrated. MCEs are represented as red open boxes, coding exons of *Rab3A* as blue blocks and non-coding UTR's as thinner blocks. The 5'UTR of *Rab3A* in ST is shown as an unfilled blue box and designated as +1. (b) A schematic representation of deletion constructs containing different upstream regions fused with pGL *Firefly Luciferase* reporter (green block). (c) *Luciferase* activity measurement of constructs in (b) in Neuro2a cells. Strong negative element(s) was defined in the first 500bp of intronic region and promoter region was defined between MCE2 and MCE4.

Figure s3. Fine mapping of the promoter region of *Rab3A* gene. (a) Gene structure of *Rab3A* is shown and the TSS of the ST is designated as +1. (b) Schematic representation of *Luciferase* constructs containing different length of region between MCE2 and 5'UTR. (c) *Luciferase* activity measurement of constructs in (b) in Neuro2a cells. Core promoter region was defined as in MCE3 region.

Figure s4. Screen of the 1.5 kb upstream region of *Rab3A* gene for potential regulatory element(s). (a) Gene structure of *Rab3A* is illustrated and the TSS of the ST is designated as +1. (b) Schematic representation of a series of *Luciferase* constructs by deleting sequence from the 3' end. (c) Measurement of *Luciferase* activity of constructs in (b) in Neuro2a cells. One enhancer was defined, corresponding to MCE1. (d) Schematic representation of a series of *Luciferase* constructs by deleting sequence from the 5' end. (e) Measurement of *Luciferase*

activity of constructs in (d) in Neuro2a cells. One enhancer was defined, corresponding to MCE2. Potential negative element(s) may exist between MCE1 and MCE2.

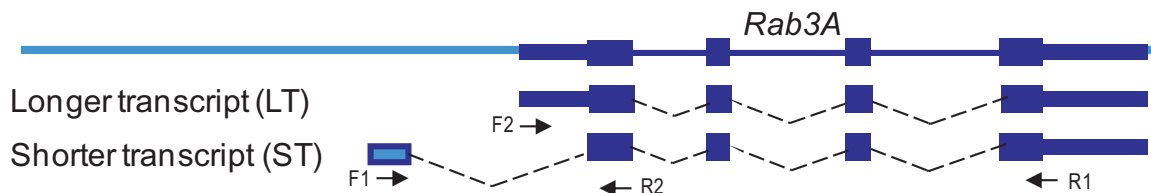
Figure s5. Screen and characterization of the first intron of *Rab3A* gene for regulatory element(s). (a) Gene structure of *Rab3A* is illustrated and the TSS of the ST is designated as +1. (b) Schematic representation of a series of *Luciferase* constructs by screening sequence of the first intron. (c) Measurement of *Luciferase* activity of constructs in (b) in Neuro2a, HEK293 and Hela cells. Two potential negative elements were identified. (d) Schematic representation of *Luciferase* constructs under control of SV40 promoter (yellow block) and SV40 enhancer (orange block). pGLc-3 was generated by placing *Rab3A* intronic sequence (+147, +556) upstream to the SV40 promoter. (e) Measurement of *Luciferase* activity of constructs in (d) in Neuro2a cells. The intronic sequence (+147, +556) when placed upstream to the SV40 promoter, fails to reduce *Luciferase* expression.

Figure s6. Mask of MCE5 effect by the presence of MCE1 and MCE2 in Neuro2a cells. (a) Gene structure of *Rab3A* is illustrated and the TSS of the ST is designated as +1. (b) Schematic representation of *Luciferase* constructs testing for MCE5. (c) *Luciferase* activity of constructs in (b) in Neuro2a, HEK293 and Hela cells. In pGLc1, MCE5 leads to decreased *Luciferase* activity in all 3 cell lines. However, in pGL-1, effect of MCE5 seems to be masked by *Rab3A* upstream region in Neuro2a cells. No such mask is observed in HEK293 and Hela cell lines. (d) *Luciferase* mRNA level of each construct was examined by RT-PCR in Neuro2a cells. β -actin gene level is used as control for RNA loading. Changes in mRNA level correspond to changes in *Luciferase* activity.

Figure s7. Statistic evaluation of the scoring strategy for MCEs. (a) Percentage of total MCEs in each expression cluster when score is more than 0. P-value was calculated based on Fisher exact test. (b) Percentage of total MCEs in each expression cluster when score is more than 90. P-value was calculated based on Fisher exact test. (c) Percentage of total MCEs in each cluster when the median's score is more than 0. The median's score represents the median of all MCEs corresponding to a given gene when arranging MCEs by gene. P-value was calculated based on Wilcoxon rank sum test.

Figure s1

a



b

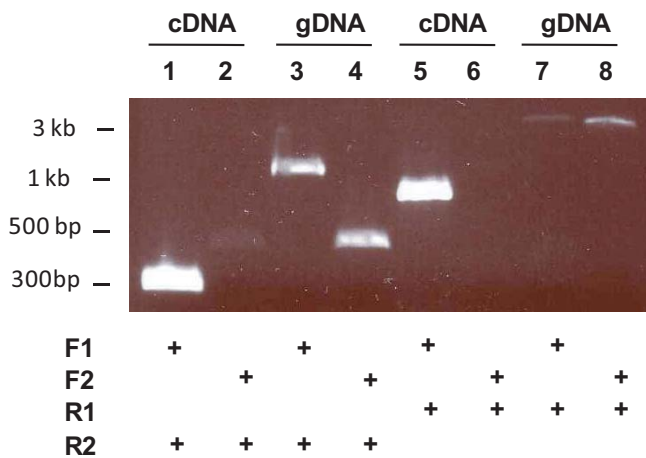
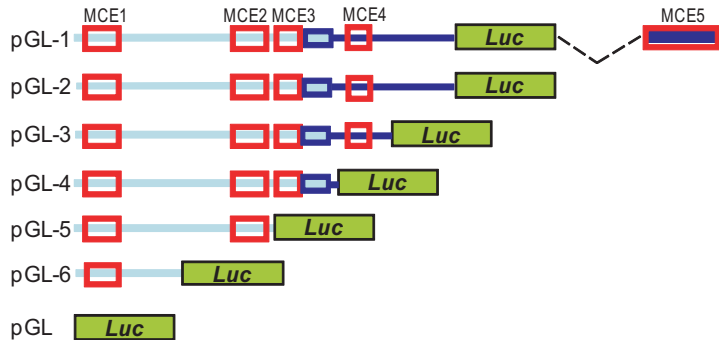


Figure s2.

a



b



c

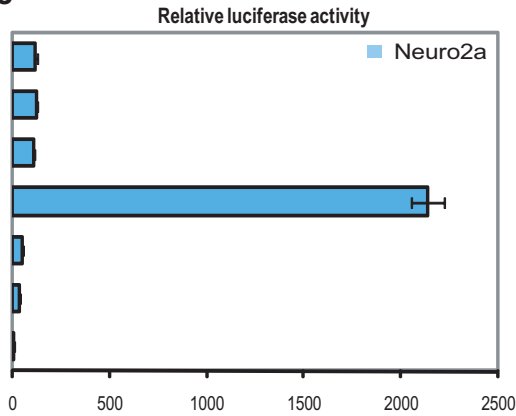
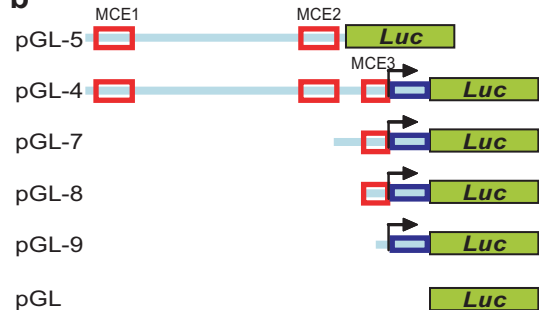


Figure s3.

a



b



c

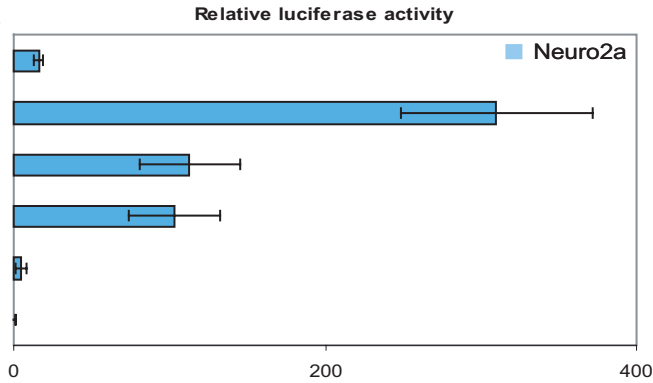


Figure s4.

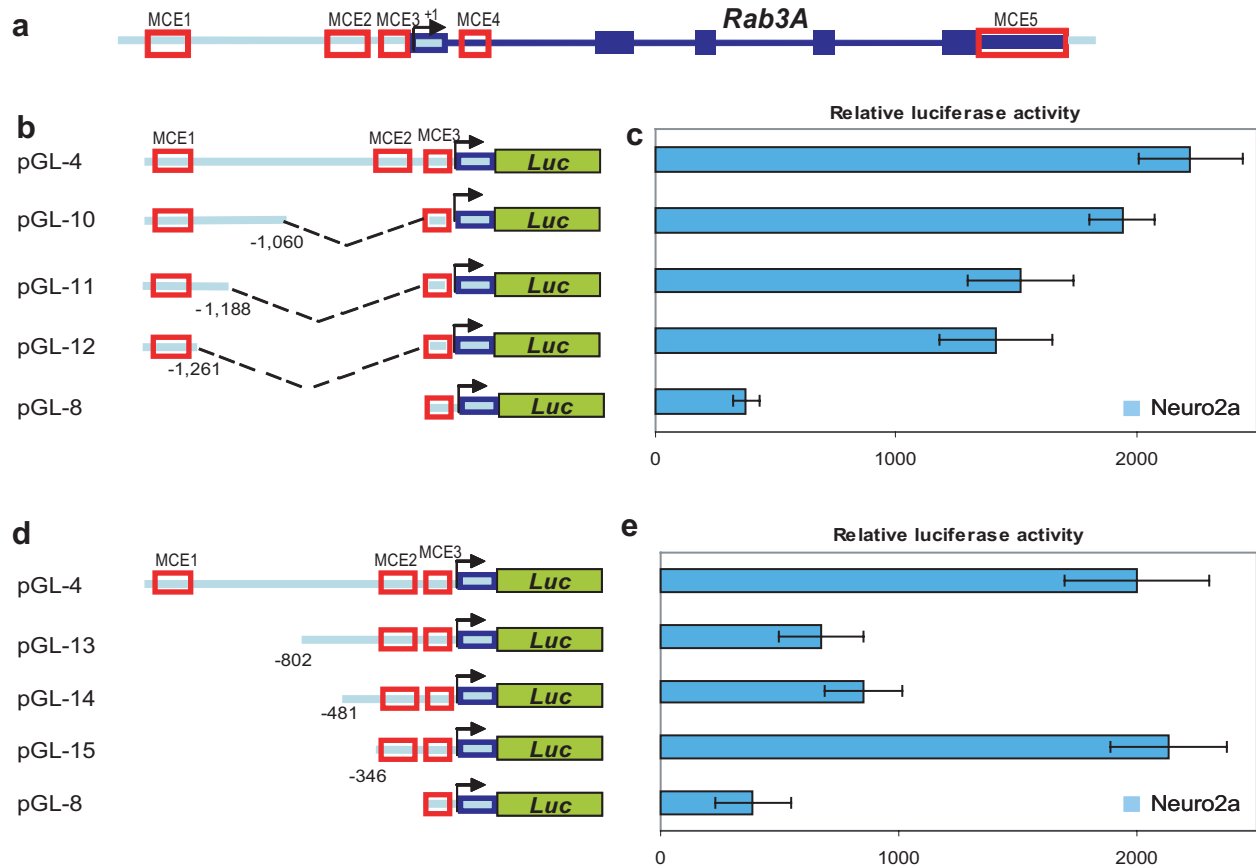


Figure s5.

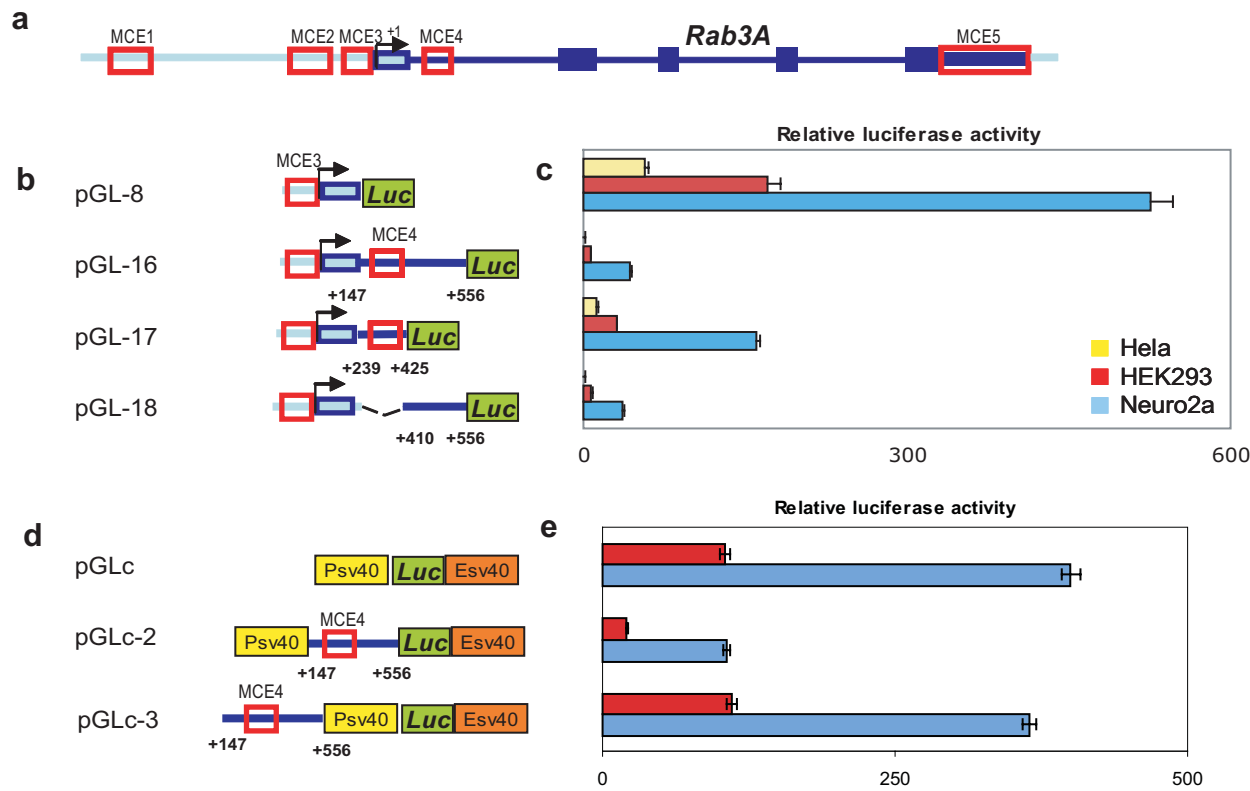
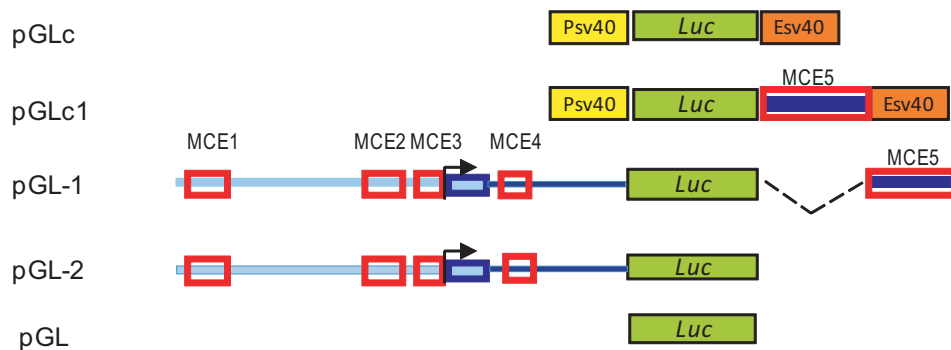


Figure s6.

a

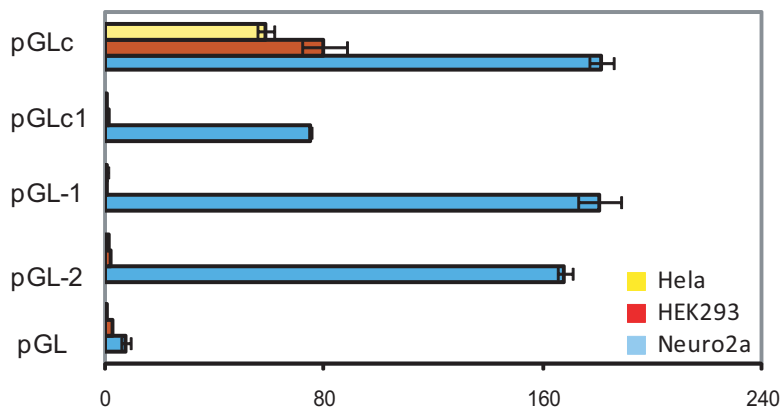


b



c

Relative luciferase activity



d RT-PCR of Luciferase mRNA

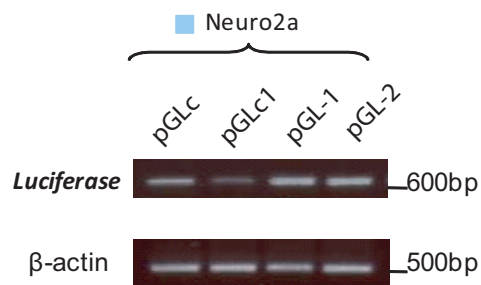


Figure s7

