

Supplementary Figure and Table Legends

Table S1. *TSC1/TSC2* variants identified in 21 normal-tumor paired SEGA samples from 20 TSC patients.

Table S2. Other somatic variants beyond *TSC1/TSC2* identified by exome sequencing analysis in 21 paired SEGA tumor-normal samples from 20 TSC subjects. Samples that are not listed had no variants.

Table S3. Previously published studies with whole exome data available for other brain tumors. These datasets were used for the mutation rate plots in Figure 2D and Figure S2.

Table S4. A de-novo clustering analysis for 3060 most variable genes identified four distinct clusters (Cluster 1-4) for the combined cohort SEGA-TCGA brain tumors-normal brain-cortical tubers-SENs-(n=745) by the consensus Bayesian non-negative matrix factorization. 15 out of 16 SEGAs were co-clustered with most GBM samples (95%, 163 out of 171) and 20% of LGG samples (103 out of 530) in Cluster 3. The set of 103 LGG in cluster 3 includes 60 astrocytomas, 15 oligoastrocytomas, 16 oligodendrogliomas and 12 unclassified brain tumors. SEN: Subependymal nodules.

Table S5-9. Differential gene expression analyses in SEGAs (n=15) compared to multiple pathological entities including: Table S5. TCGA astrocytomas (n=60), Table S6. TCGA oligodendrogliomas (n=16), Table S7. TCGA glioblastomas (n=153), Table S8. gangliogliomas (n=10), Table S9. Cortical tubers (n=15). All upregulated/downregulated genes for adjusted p-value/FDR<0.05 are presented. Expression RSEM values are shown in log₂ scale.

Table S10. A total of 190 genes were differentially expressed in each of the 5 comparisons. Fold change for each comparison is shown, as well as median fold change from all 5 comparisons. Sorted by median fold change.

Table S11. Differential gene expression analyses in SEGA (n=15) compared to normal brain samples (n=11) [6]. All upregulated/downregulated genes for adjusted p-value/FDR<0.05 are presented. Expression FPKM values are shown in log₂ scale. Notably, 151 of these 190 common DEGs were also differentially expressed in SEGA in comparison to normal brain (highlighted in orange).

Table S12. Pathway analysis revealed a set of 18 GO gene sets that were enriched for the 190 common DEGs by GSEA; 8 GO sets were enriched for the 68 upregulated DEGs, top; 10 were enriched for 122 downregulated DEGs bottom.

Table S13. IHC staining results for the top differentially expressed TFs in SEGAs compared to other brain tumors and normal brain as a control for comparison. The table includes information for each marker and sample stained (age, gender, diagnosis). At least 3 different samples were stained for each tumor type and normal brain (infant or/and young control below <20 years). The extent of immunoreactivity was graded according to the percentage of cells with positive staining (score **0**: 0-5%; **1+**: 5-25%; **2+**: 25-50%; **3+**: 50-75%; or **4+**: 75-100%) and the intensity of the staining (weak, moderate, or strong). Positivity was defined as moderate or strong nuclear or cytoplasmic immunoreactivity.

For SIX3, positivity was defined as cytoplasmic immunoreactivity of moderate or strong intensity in at least 25% of cells. Cytoplasmic positivity for SIX3 was observed in SEGA (4/5, 80%), the balloon cells of cortical tubers (3/3, 100%), PXA, (2/3, 67%) and a

single glioblastoma (1/3, 33%). No appreciable staining was present in low grade astrocytomas (0/3), oligodendrogliomas (0/3), gangliogliomas (0/3), subependymomas (0/3), or normal brain (0/5).

For IRF6, positivity was defined as cytoplasmic immunoreactivity of any intensity in at least 5% of cells. IRF6 was the most specific marker, showing cytoplasmic positivity only in SEGAs (3/3, 100%) and the balloon cells of cortical tubers (2/3, 67%). No staining was detected in subependymomas (0/3), gangliogliomas (0/3), low grade astrocytomas (0/2), oligodendrogliomas (0/4), glioblastomas (0/3), PXAs (0/3), or normal brain sections (0/4).

For VAX1, positivity was defined as nuclear immunoreactivity of any intensity in at least 5% of cells. Nuclear VAX1 immunopositivity was observed in SEGAs (2/2, 100%), cortical tubers (3/3, 100%), oligodendrogliomas (2/4, 50%), subependymomas (2/3, 66%), ganglioglioma (1/3, 33%), PXA (1/3, 33%), and in both glia and neurons of normal infant brains (5/5, 100%); and hence was not specific to SEGA. No staining was seen in low-grade astrocytomas (0/4) or glioblastomas (0/3).

For ZBTB20, positivity was defined as nuclear immunoreactivity of moderate or strong intensity in 25% or greater of cells. ZBTB20 demonstrated the least specificity, showing strong nuclear staining in all tumors and normal.

We also performed IHC for CTSK (mouse monoclonal, Abcam, cat. no. ab37259, dilution 1:5000); HCRTR2 (rabbit polyclonal, Atlas antibodies, Cat. No. HPA054516, dilution 1:100), EOMES (rabbit polyclonal, Atlas Antibodies, HPA028896, dilution 1:50), TWIST2 (rabbit polyclonal, Invitrogen, PA5-78211, dilution 1:100), and GPNMB (rabbit polyclonal, Cell Signaling, Cat. No 38313S, dilution 1:5000).

For HCRTR2, positivity was defined as cytoplasmic immunoreactivity of any intensity in at least 5% of cells. HCRTR2 cytoplasmic immunoreactivity was observed in SEGAs (3/3, 100%), gangliogliomas (3/4, 75%), and a single subependymoma (1/3, 33%). No staining

was detected in low grade astrocytomas (0/3), oligodendrogliomas (0/4), PXAs (0/3), glioblastomas (0/3), or normal brain sections (0/4). Interestingly, this was the only marker not to show positivity in cortical tubers.

For CTSK, positivity was defined as cytoplasmic immunoreactivity of moderate or strong intensity in at least 50% of cells. CTSK similarly showed strong cytoplasmic positivity in SEGA (5/5, 100%), while less frequent staining was seen in cortical tubers (1/3, 33%) and glioblastomas (1/3, 33%). No staining was detected in subependymomas (0/3), gangliogliomas (0/3), low grade astrocytomas (0/3), oligodendrogliomas (0/4), PXAs (0/3), or normal brain sections (0/4).

The IHC for TWIST2 and EOMES did not show specificity in SEGAs (data not shown).

Table S14. WGCNA analysis revealed a set of 41 different modules (ME, clusters of co-expressed genes) in SEGA in comparison to TCGA low grade gliomas, glioblastomas, gangliogliomas and cortical tubers. Higher values for correlation scores for modules mean that genes in the same module are likely to be upregulated in SEGA and downregulated in the other brain tumors, and the reverse applies for more negative values.

Table S15. WGCNA analysis showing all modules, and genes belonging to each. Genes with higher kME are thought to be "hub" genes that are involved in regulating expression within that module. Genes with positive values means that they are upregulated in SEGAs, whereas negative values indicate downregulation in SEGAs.

Fig. S1. Allele ratio plots to assess Copy Neutral Loss of Heterozygosity (CN-LOH) on chromosomes 9 and 16 for 21 SEGAs analyzed by exome sequencing.

Fig. S2. Somatic mutation burden of SEGAs in comparison to a variety of TCGA and non-TCGA brain tumors and other cancers.

Fig. S3a-e. Volcano plots showing all differentially expressed genes (DEGs, top) and all differentially expressed TFs (bottom) in SEGA versus individual other brain tumors and cortical tubers (FDR<0.05). The x-axis shows \log_2 of the fold change in expression level, and the y-axis shows $-\log_{10}$ of the p-value for statistical significance of the difference. Each dot represents a different gene, blue denotes \log_2 (fold change) > -1 and < 1; green denotes \log_2 (fold change) > 1, up-regulated genes; red denotes \log_2 (fold change) < -1, down-regulated.

f. Heatmap showing clustering of SEGA samples among other brain samples (columns) using the 190 DEGs.

Fig. S4. Box plots for housekeeping genes show similar levels of gene expression across SEGA tumors and other tumors/cortical tubers. P-values are indicated, and reflect minimal to no major differences in expression level.

Fig. S5. Box plots for the very top DEG (HCRTR2) as well as other top differentially expressed (upregulated) TFs (TWIST2, EHF, ZBTB20) in SEGAs compared to TCGA tumors (2463 tumors of 27 different histologic types), gangliogliomas and cortical tubers. CTSK, highly expressed in TSC associated kidney angiomyolipoma is also highly expressed in SEGAs. Similar high expression pattern for GPNMB, a known specific marker for SEGAs. Gene expression is shown in RSEM values.

Fig. S6. Box plots for the very top DEG (HCRTR2) as well as other top differentially expressed (upregulated) TFs (TWIST2, EHF, ZBTB20) in SEGAs compared to GTEx human

normal tissues (~8,500 samples from 30 normal tissue types, v6p release). ZBTB20, HCRTR2 and GPNMB are very highly expressed in SEGAs, whereas expression TWIST2, EHF and CTSK is similar to other normal tissues. Gene expression is shown in FPKM values.

Fig. S7. IHC images for protein markers CTSK, HCRTR2, ZBTB20 and GPNMB.

Scale bar: 400 μ m

Fig. S8. All 41 distinct modules of highly inter-correlating transcripts/ co-expressed genes in SEGA, TCGA brain tumors (Cluster 3 from Figure 3A), gangliogliomas and cortical tubers. Each color indicates a different tumor/sample type. Red: SEGA, pink: TCGA astrocytomas and oligodendrogliomas, green: cortical tubers, blue: TCGA glioblastomas, purple: gangliogliomas. X axis represents sample type and Y axis shows relative correlation (positive or negative; log scale).

No LOH of chromosome 9 in SEGA S1

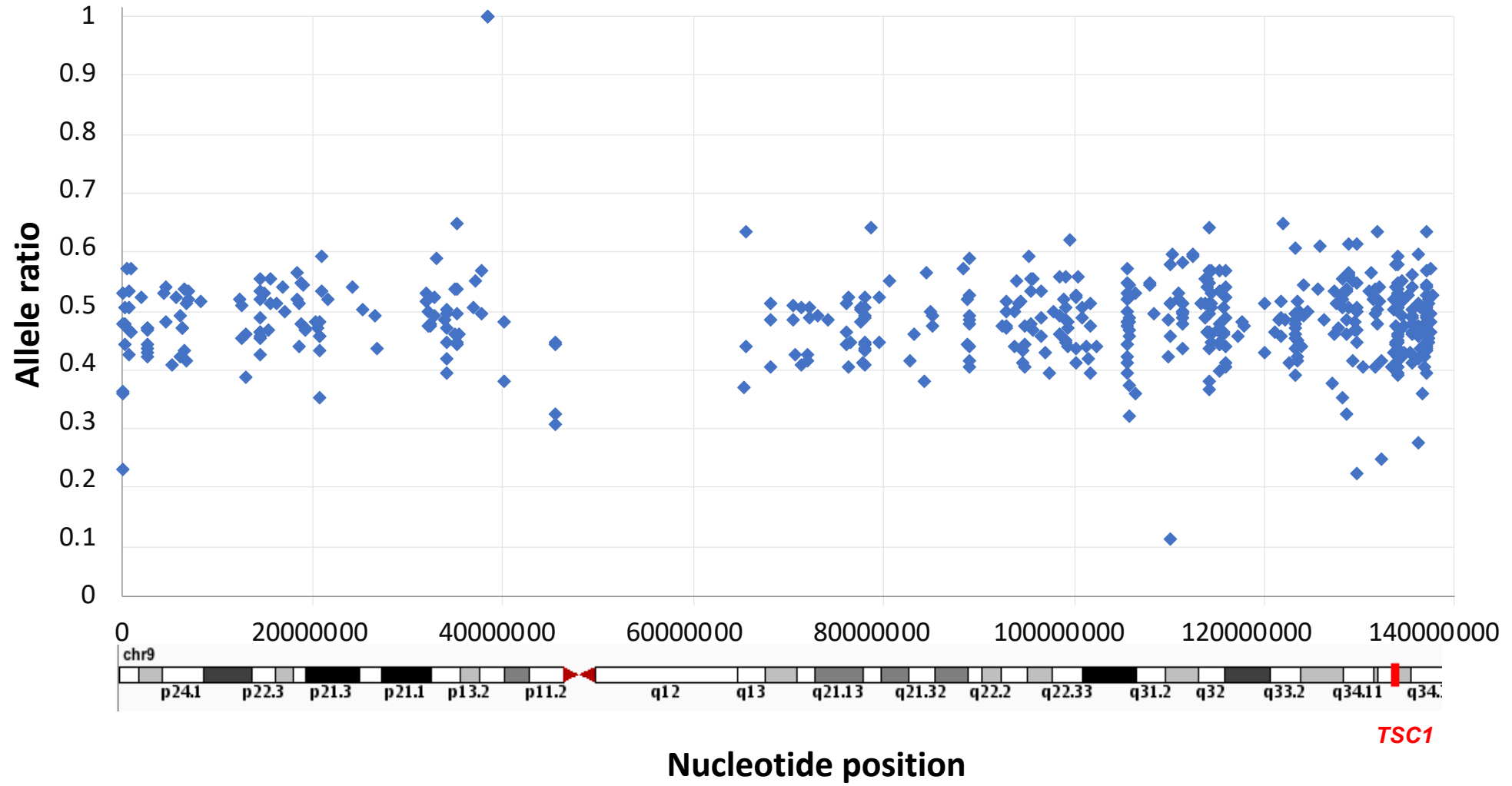
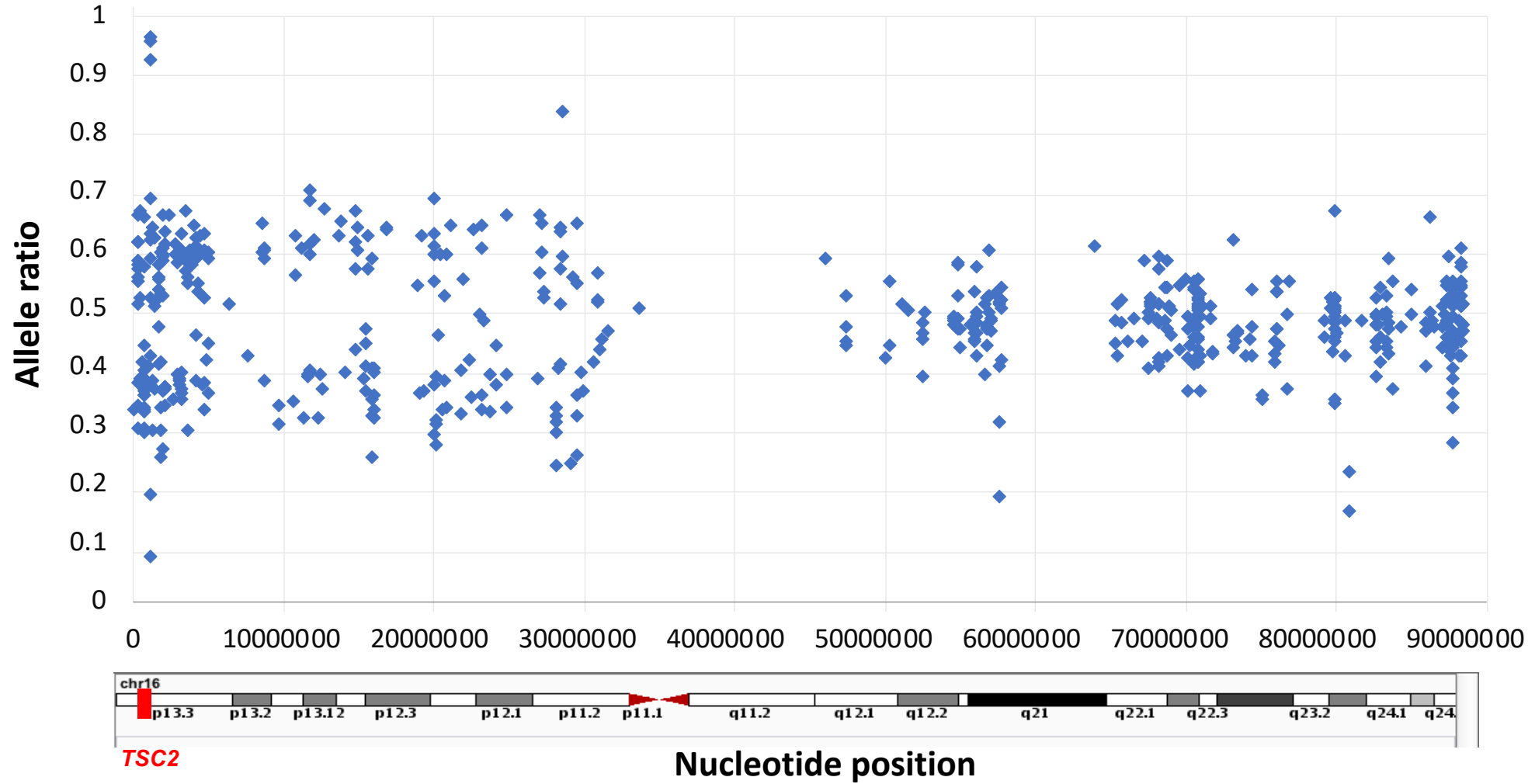
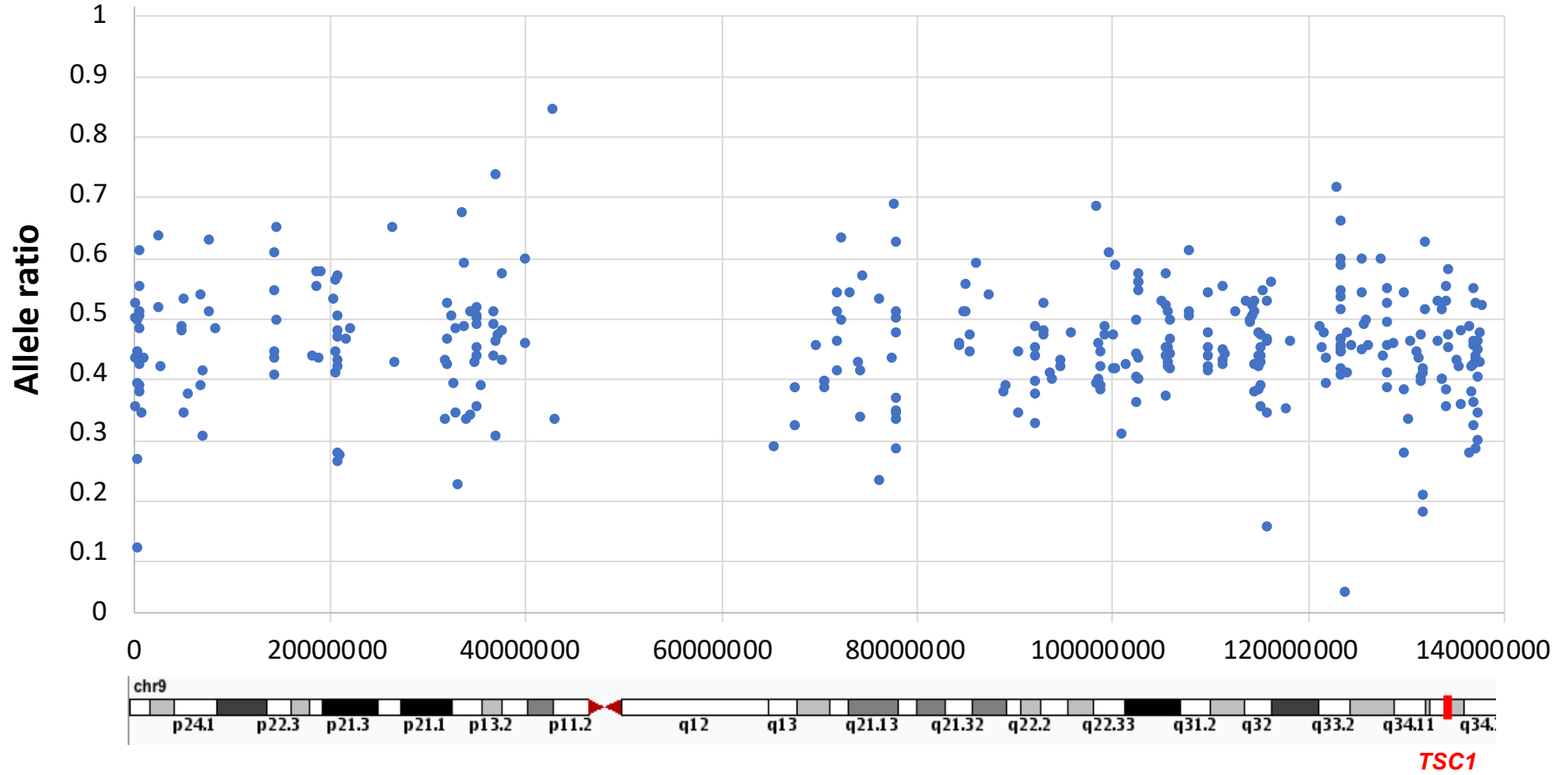


Fig. S1

LOH mapping of chromosome 16 in SEGA S1



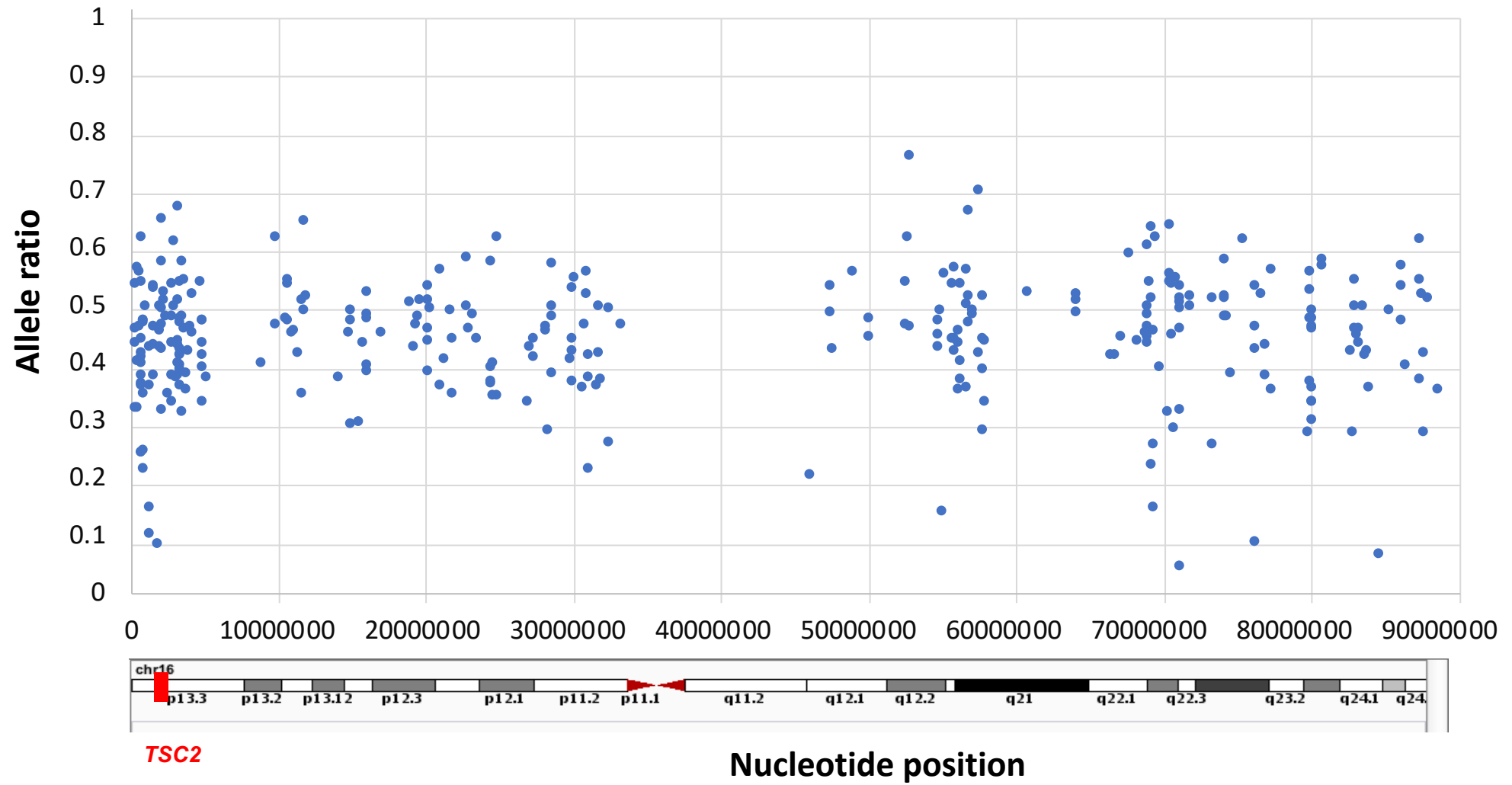
No LOH of chromosome 9 in SEGA S2



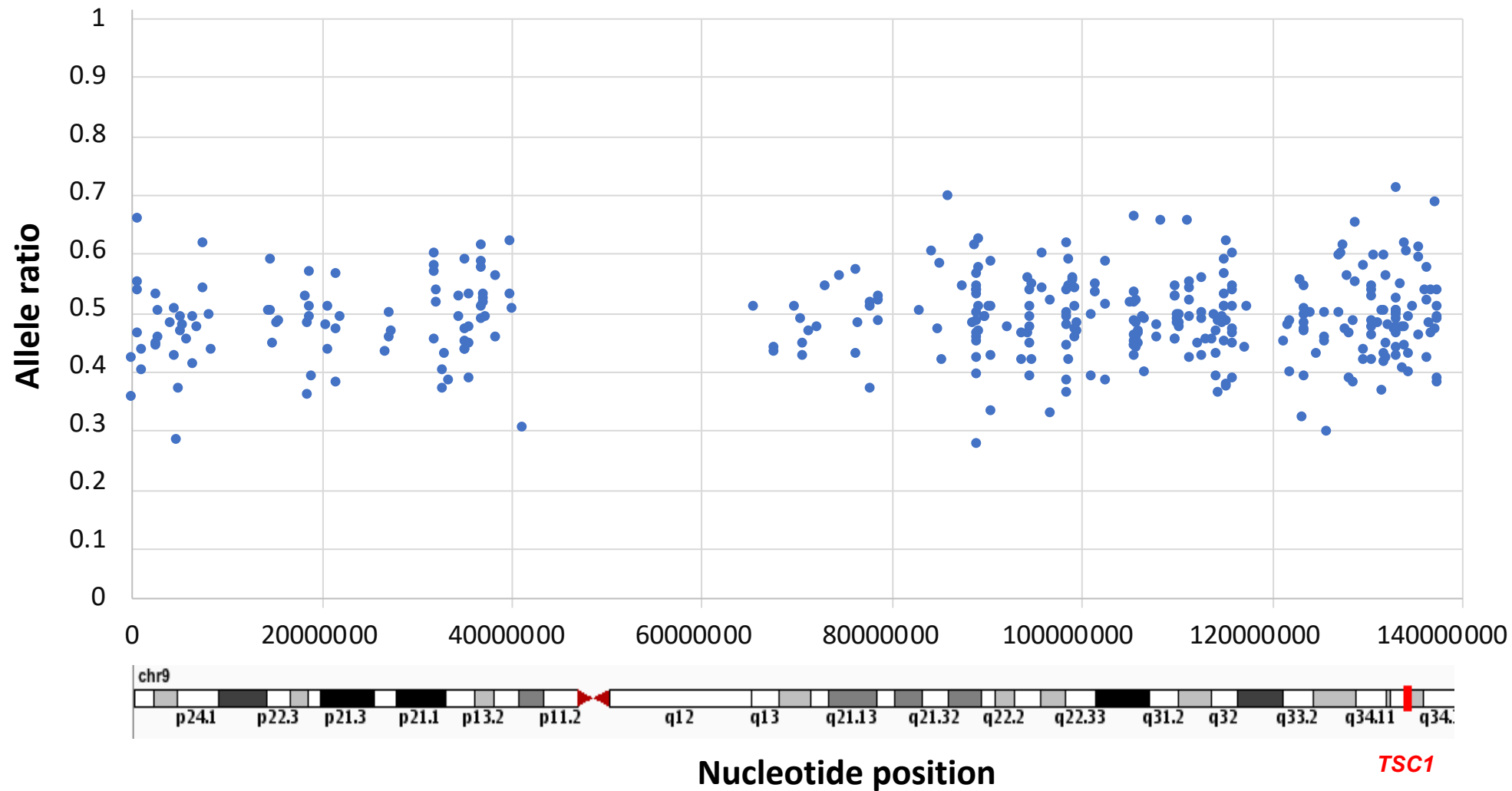
Nucleotide position

TSC1

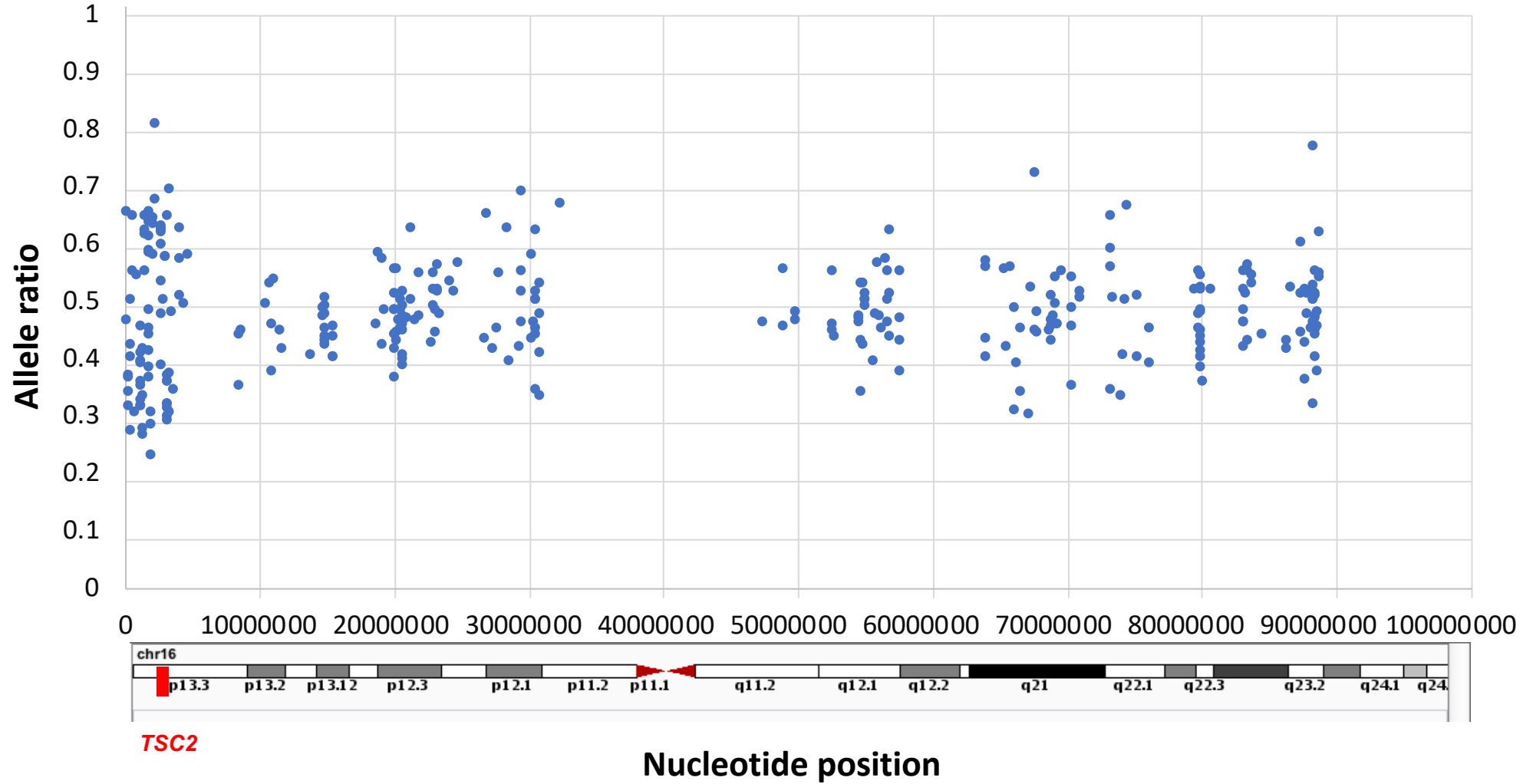
No LOH of chromosome 16 in SEGA S2



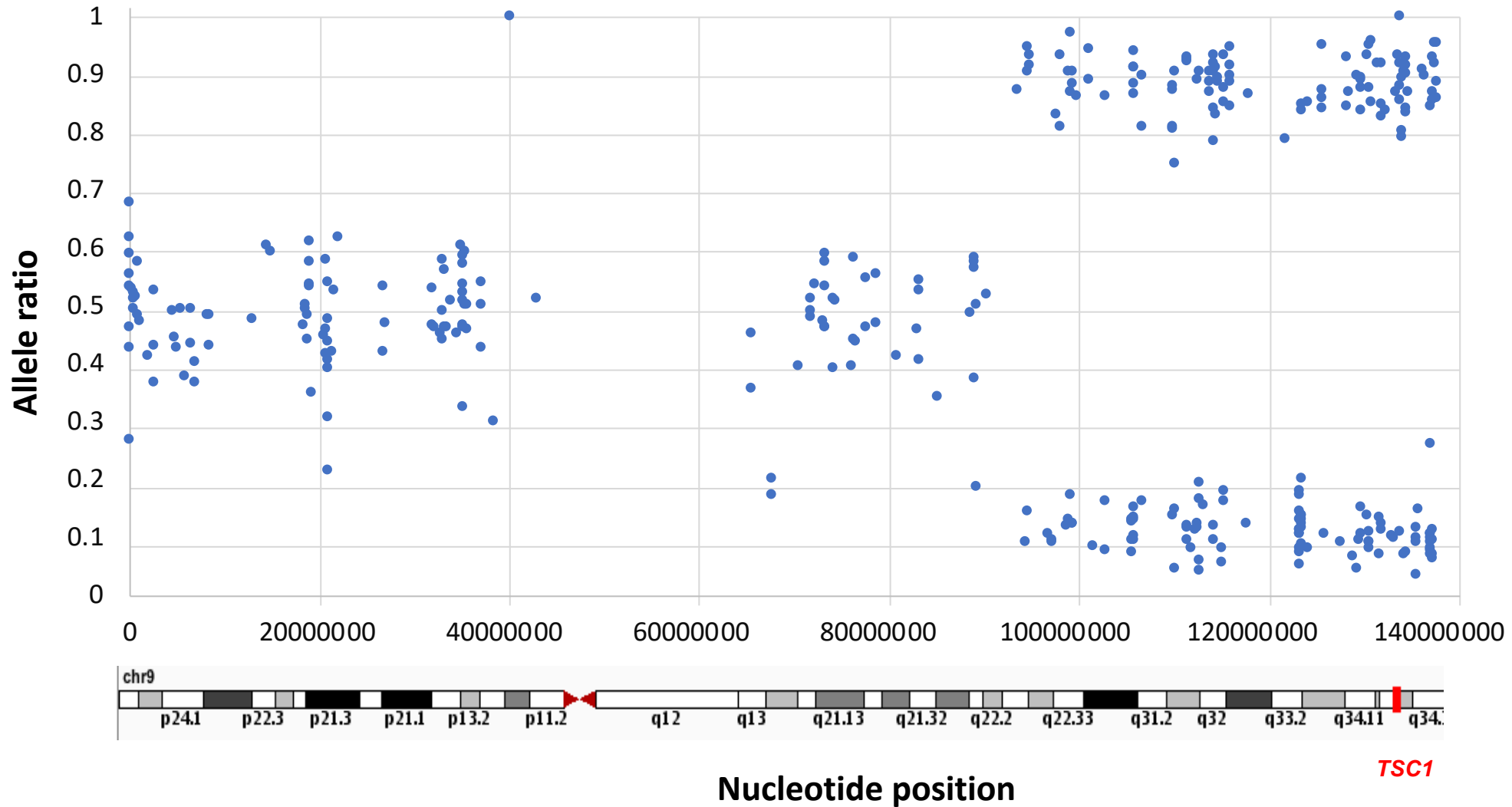
No LOH of chromosome 9 in SEGA S3



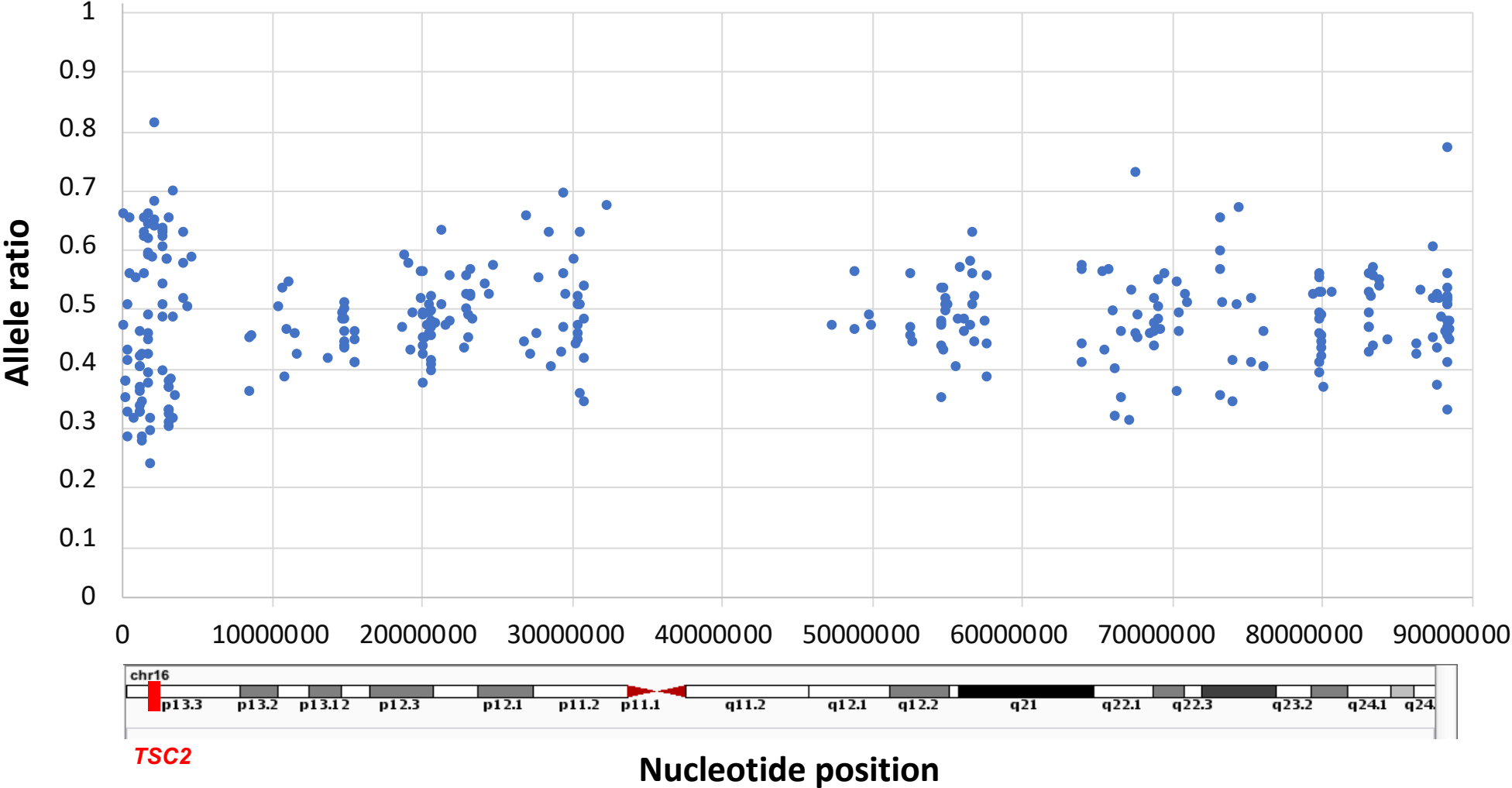
LOH mapping of chromosome 16 in SEGA S3



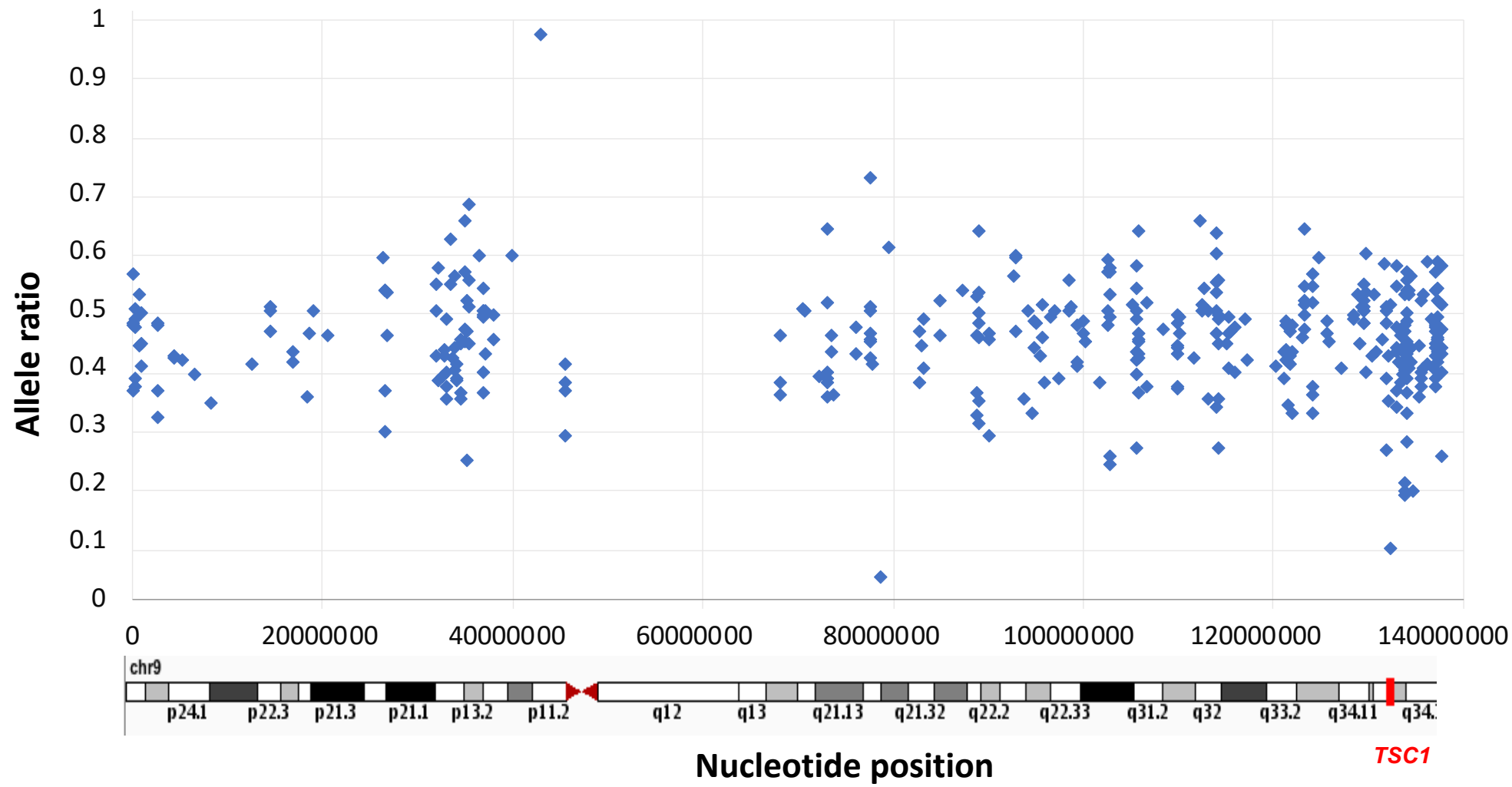
LOH mapping of chromosome 9 in SEGA S4



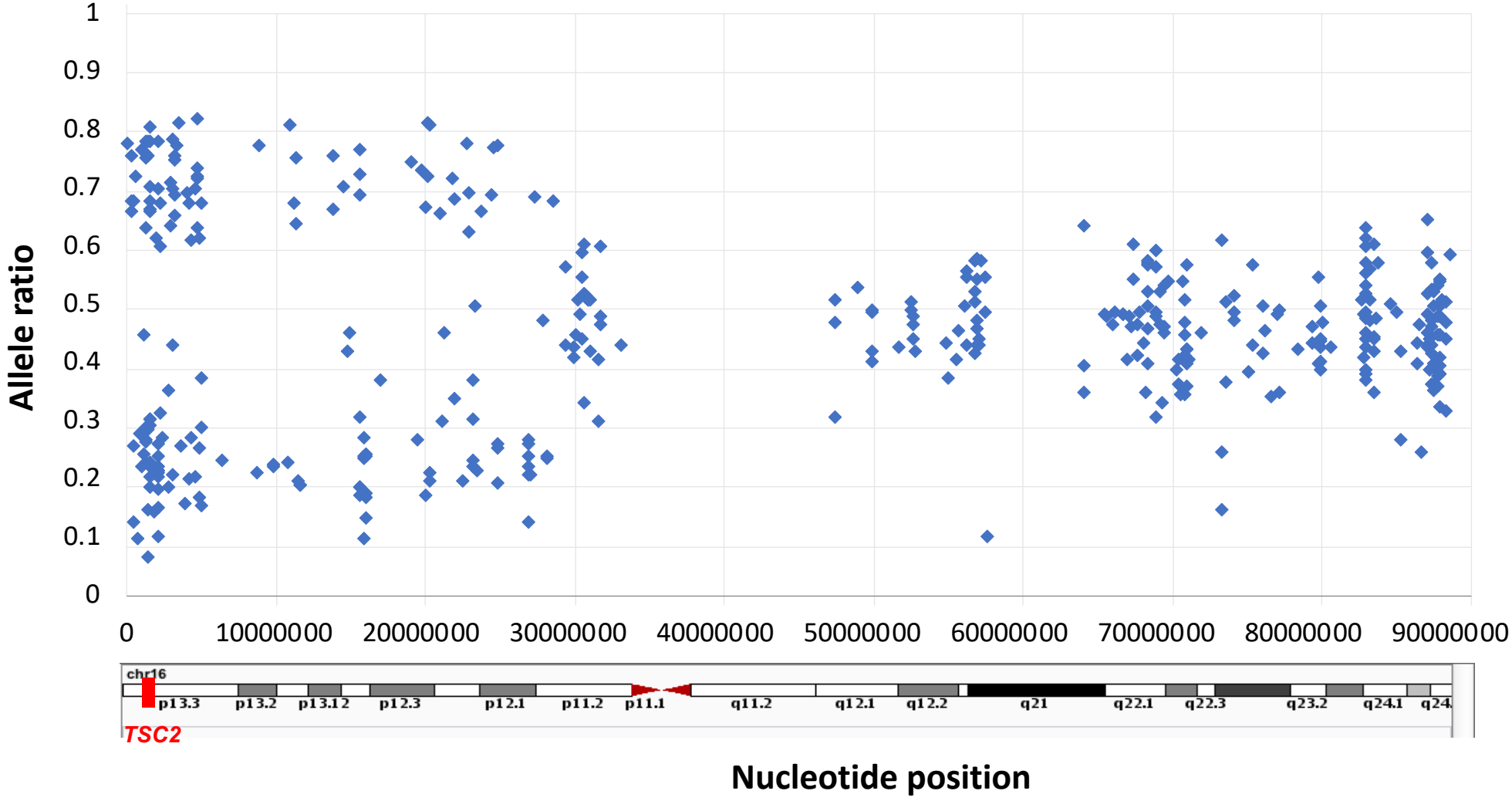
No LOH of chromosome 16 in SEGA S4



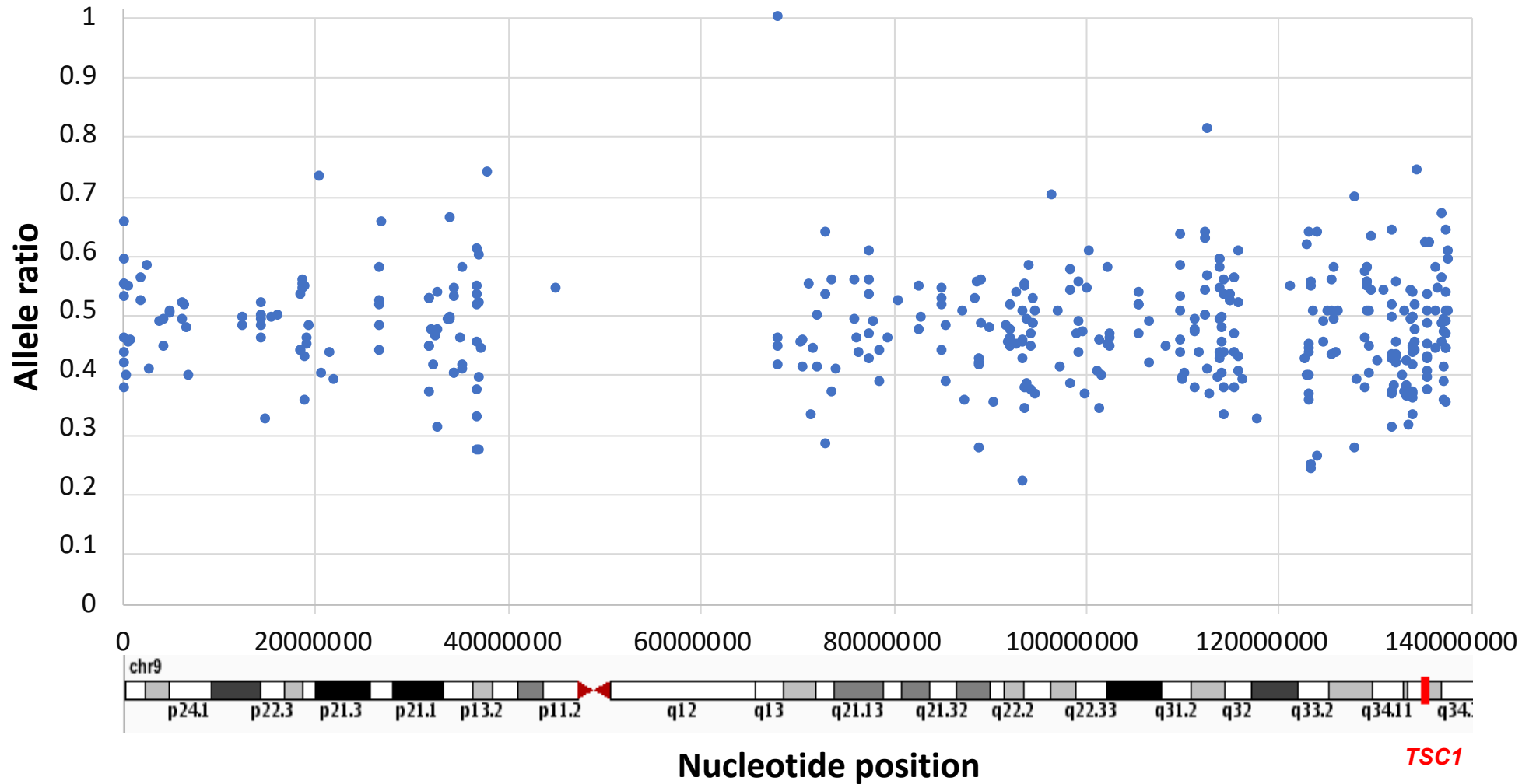
No LOH of chromosome 9 in SEGA S5



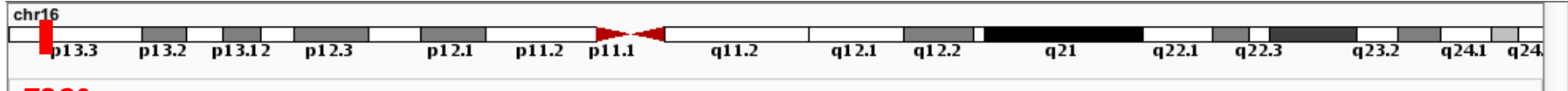
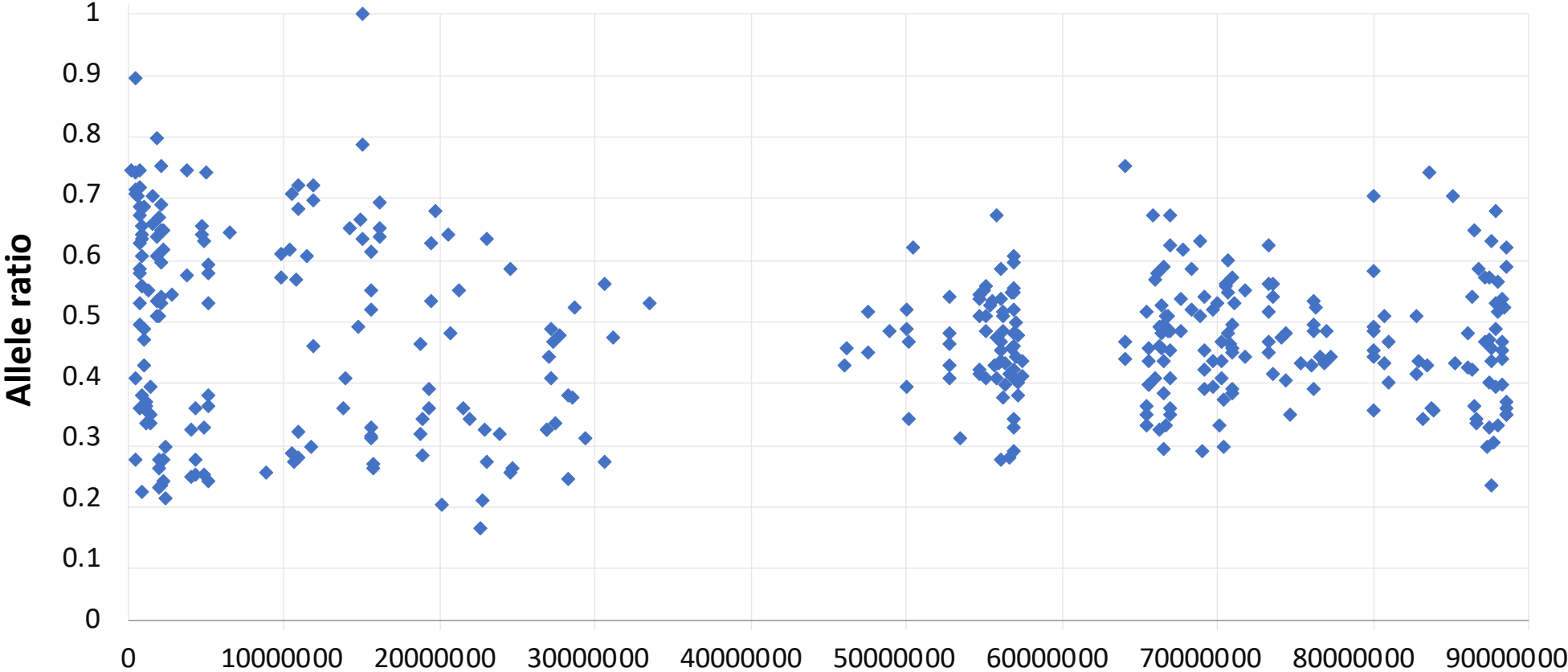
LOH mapping of chromosome 16 in SEGA S5



No LOH of chromosome 9 in SEGA S6



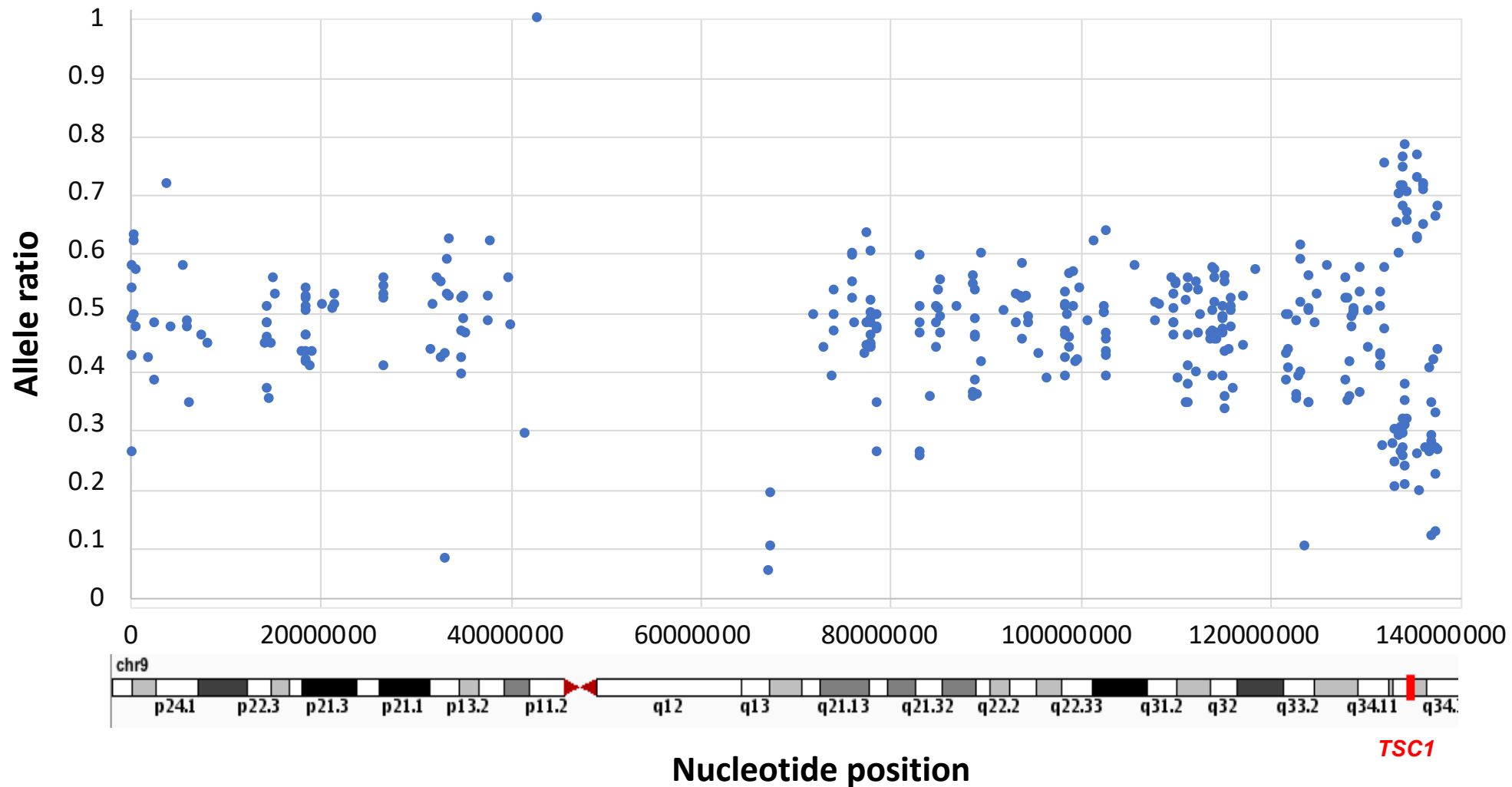
LOH mapping of chromosome 16 in SEGA S6



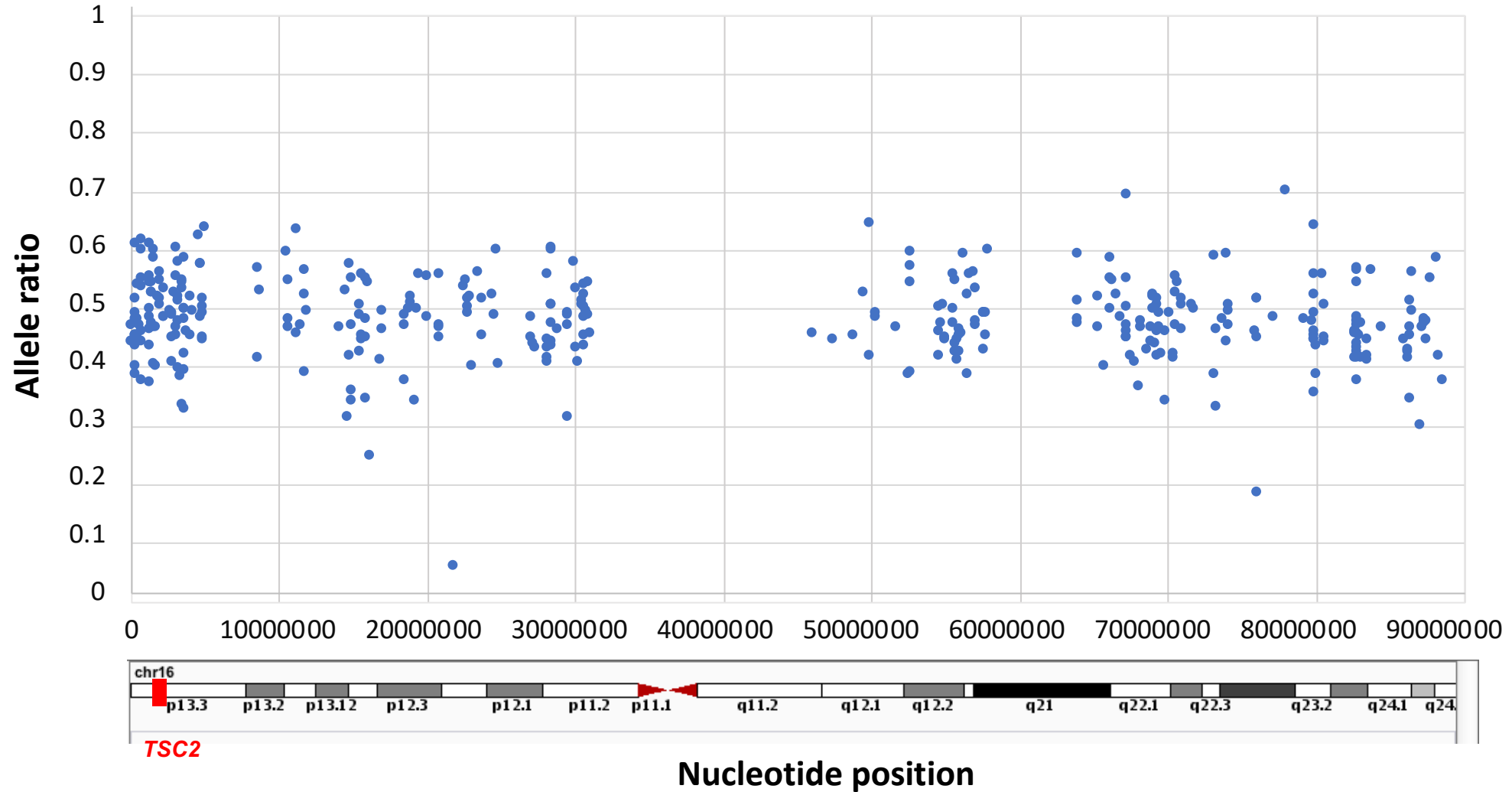
TSC2

Nucleotide position

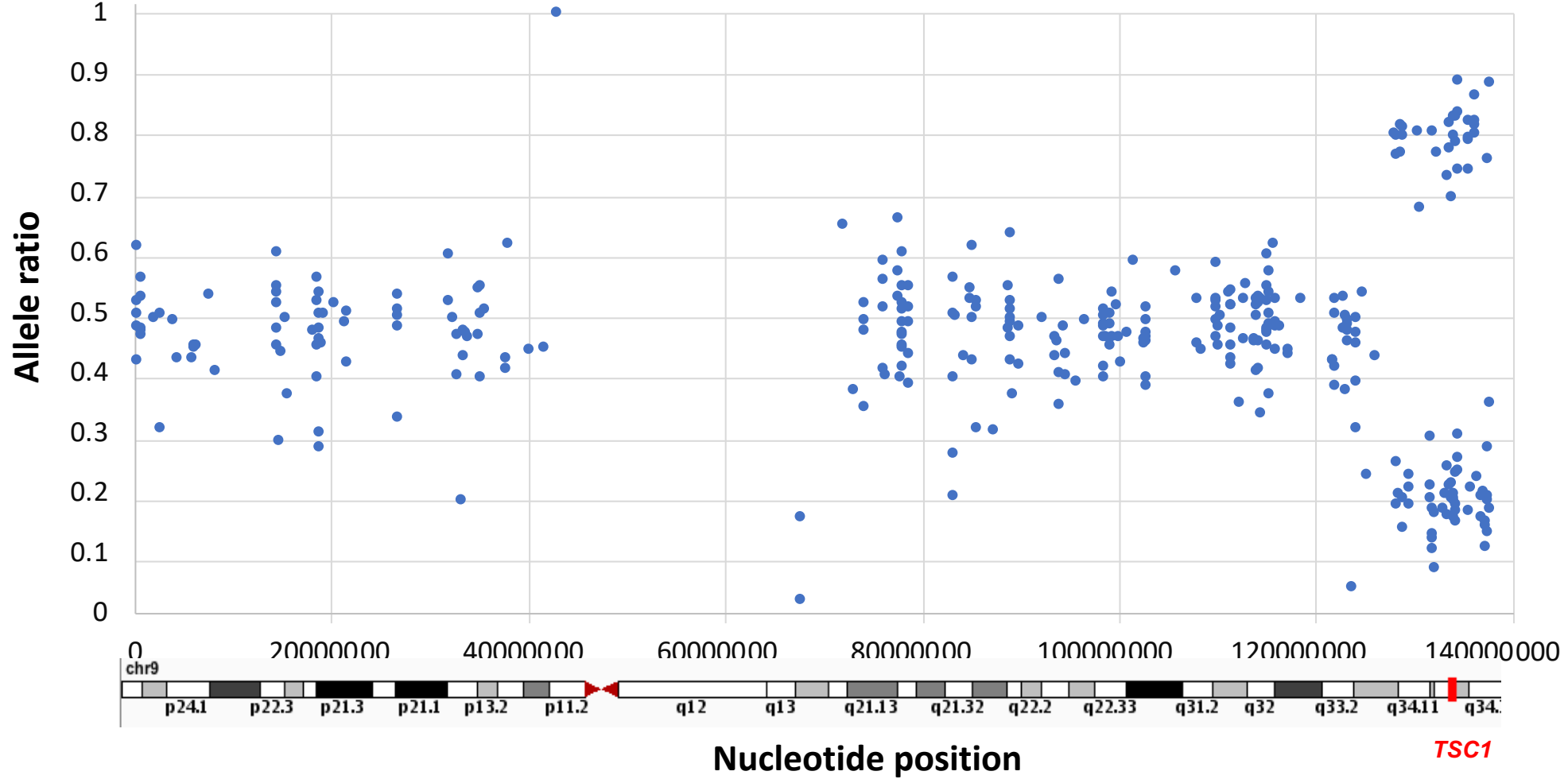
LOH mapping of chromosome 9 in SEGA S7a



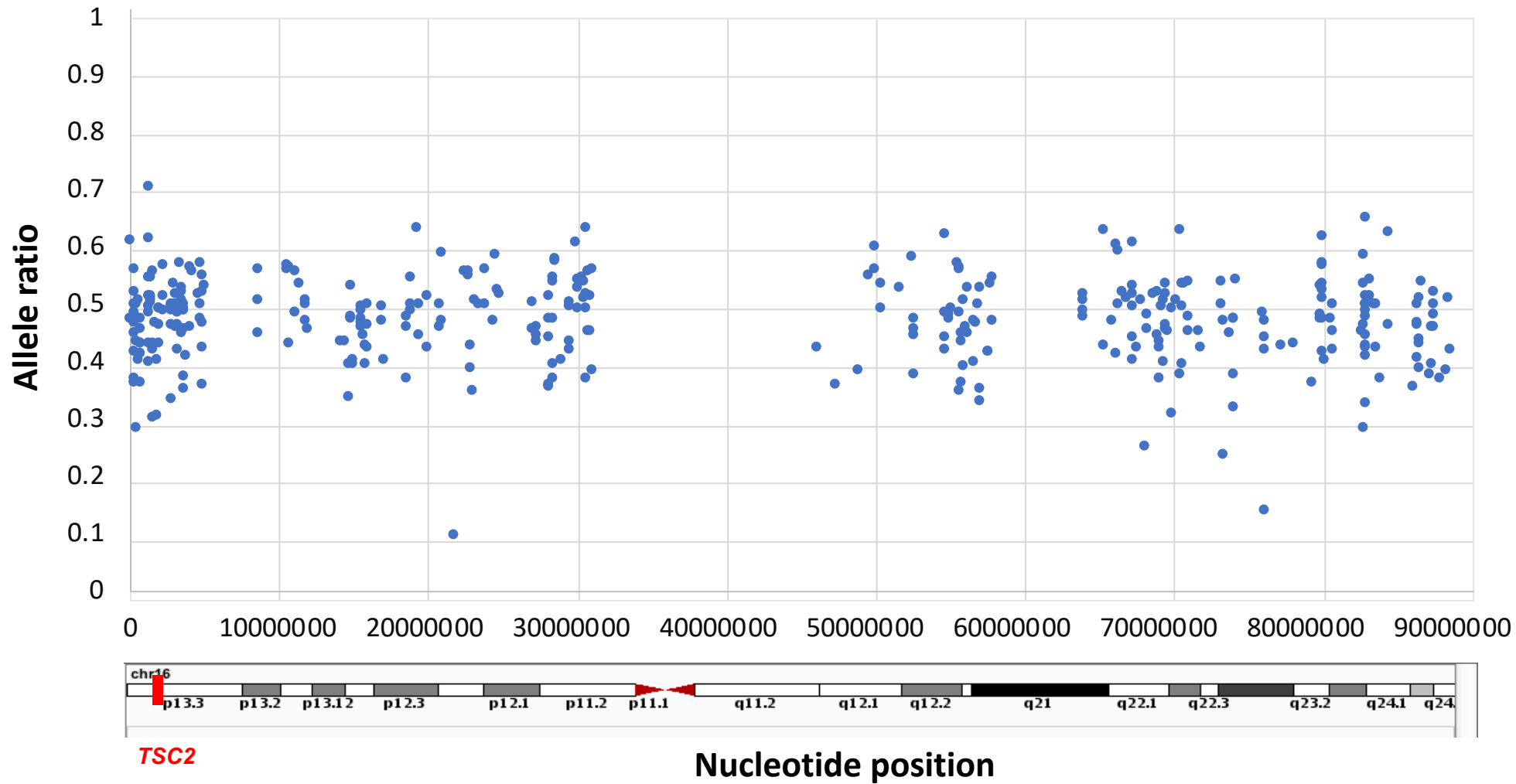
No LOH of chromosome 16 in SEGA S7a



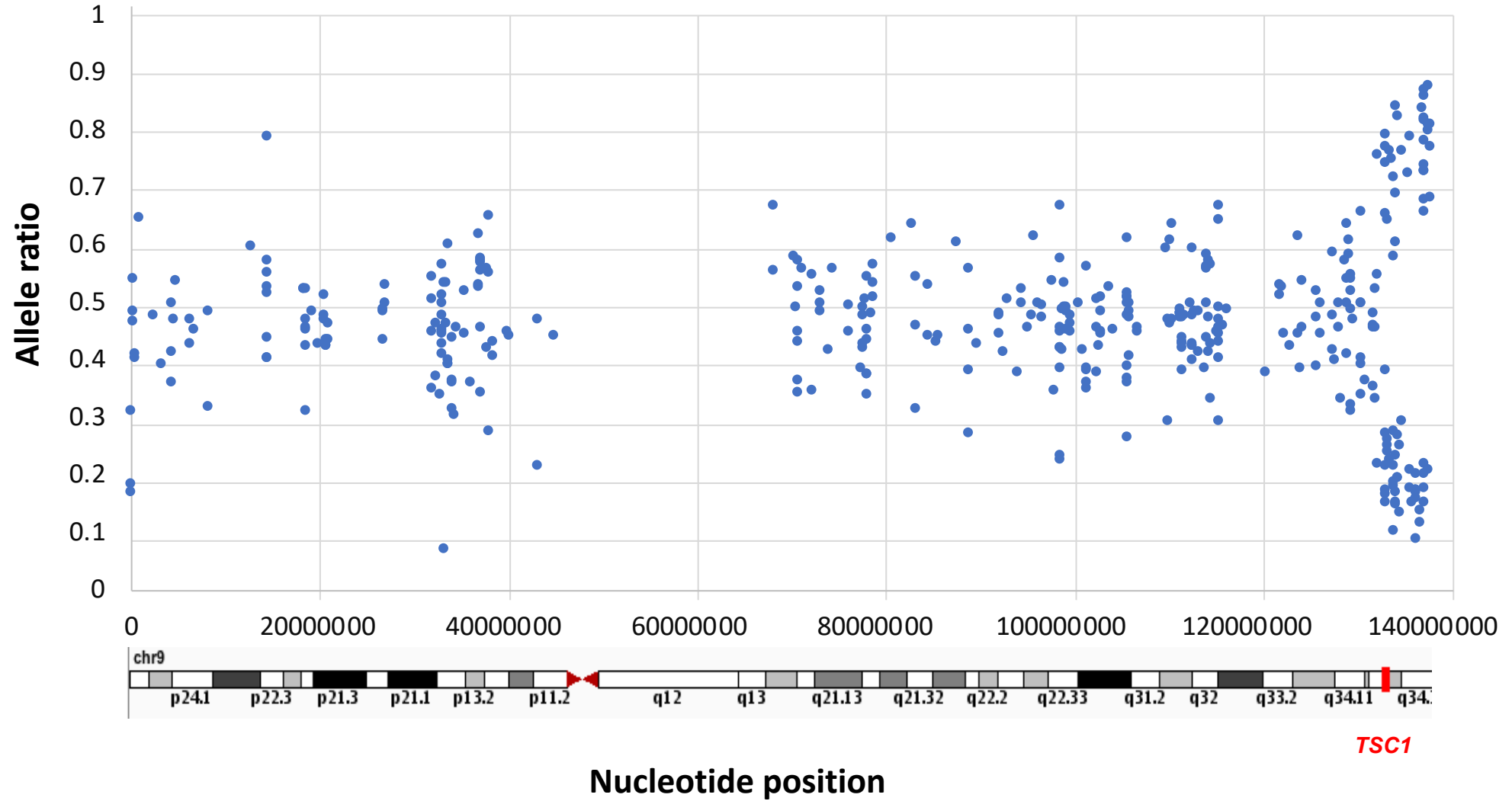
LOH mapping of chromosome 9 in SEGA S7b



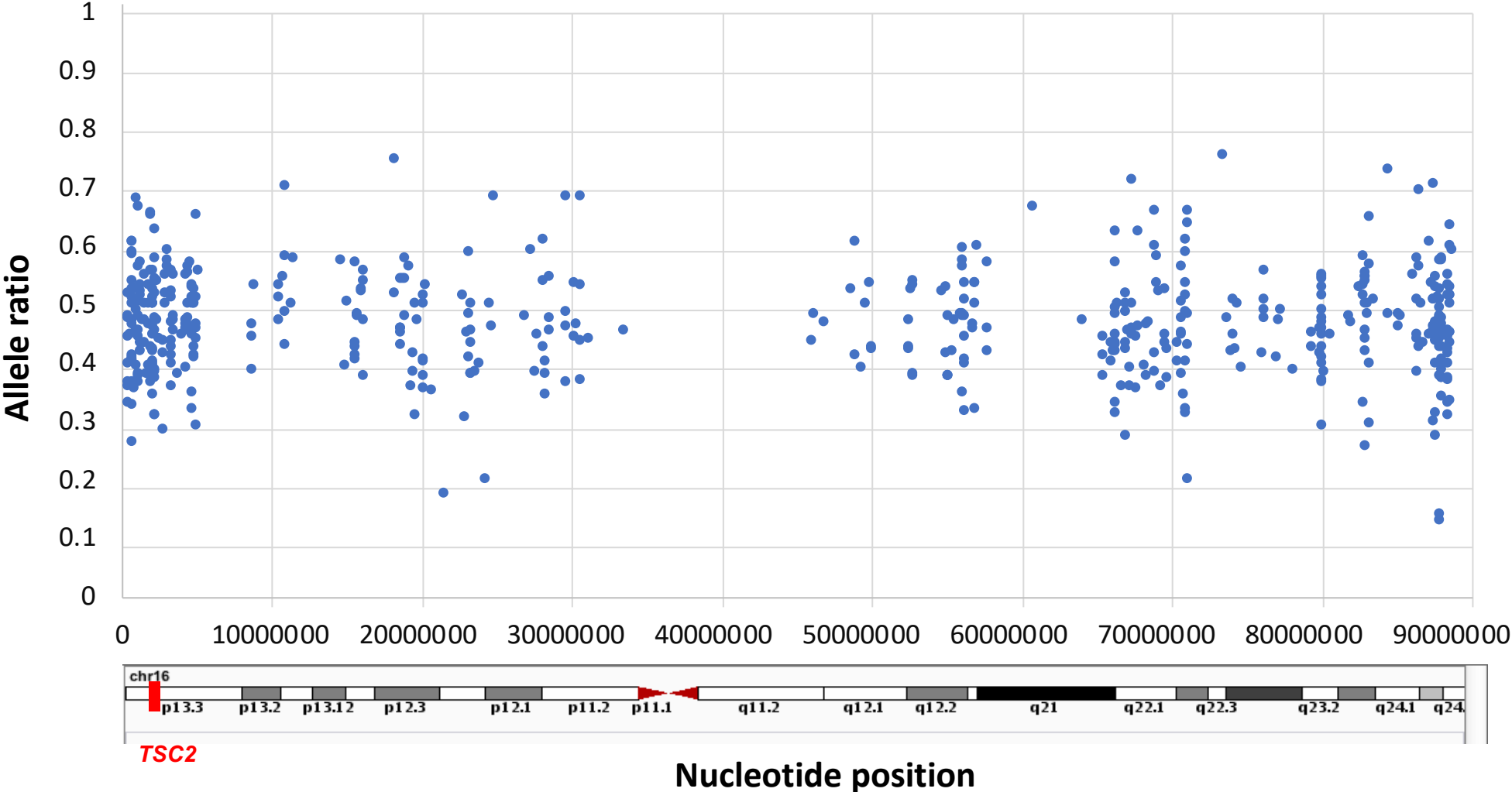
No LOH of chromosome 16 in SEGA S7b



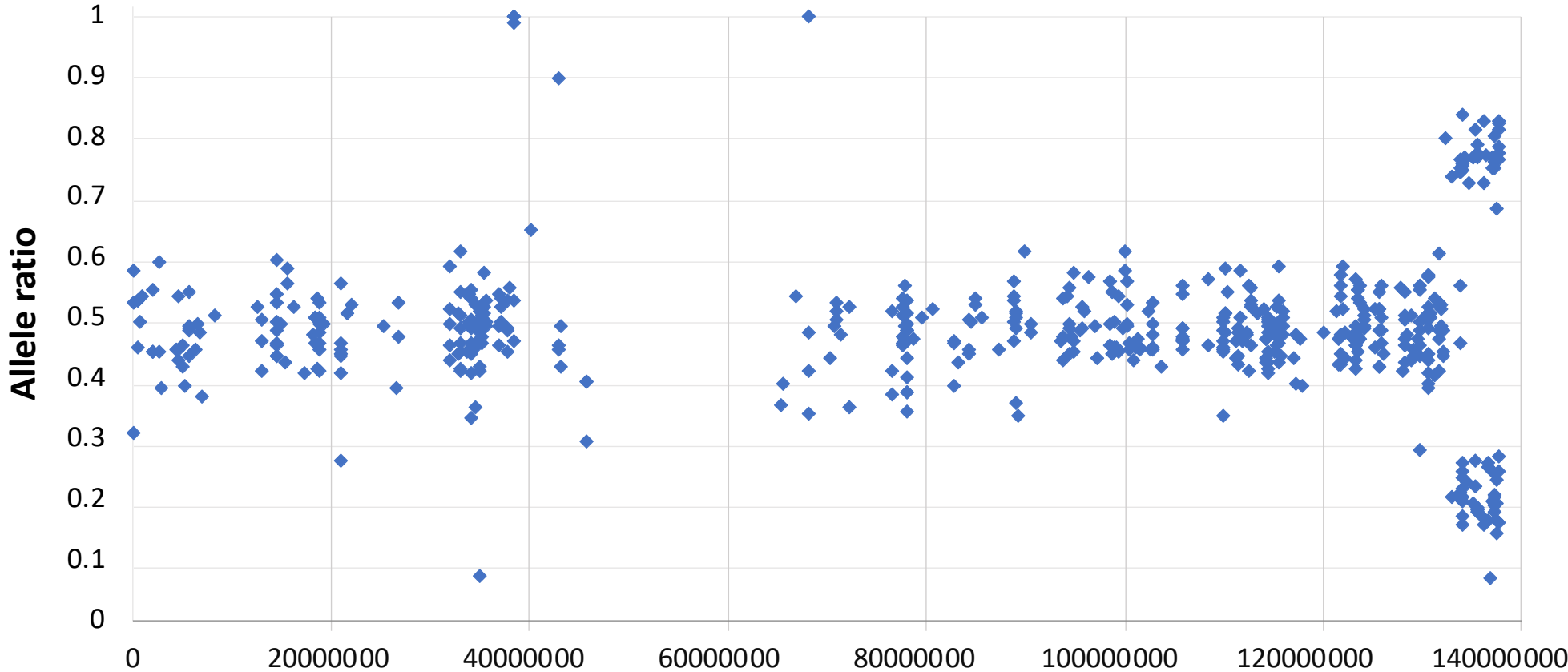
LOH mapping of chromosome 9 in SEGA S8



No LOH of chromosome 16 in SEGA S8



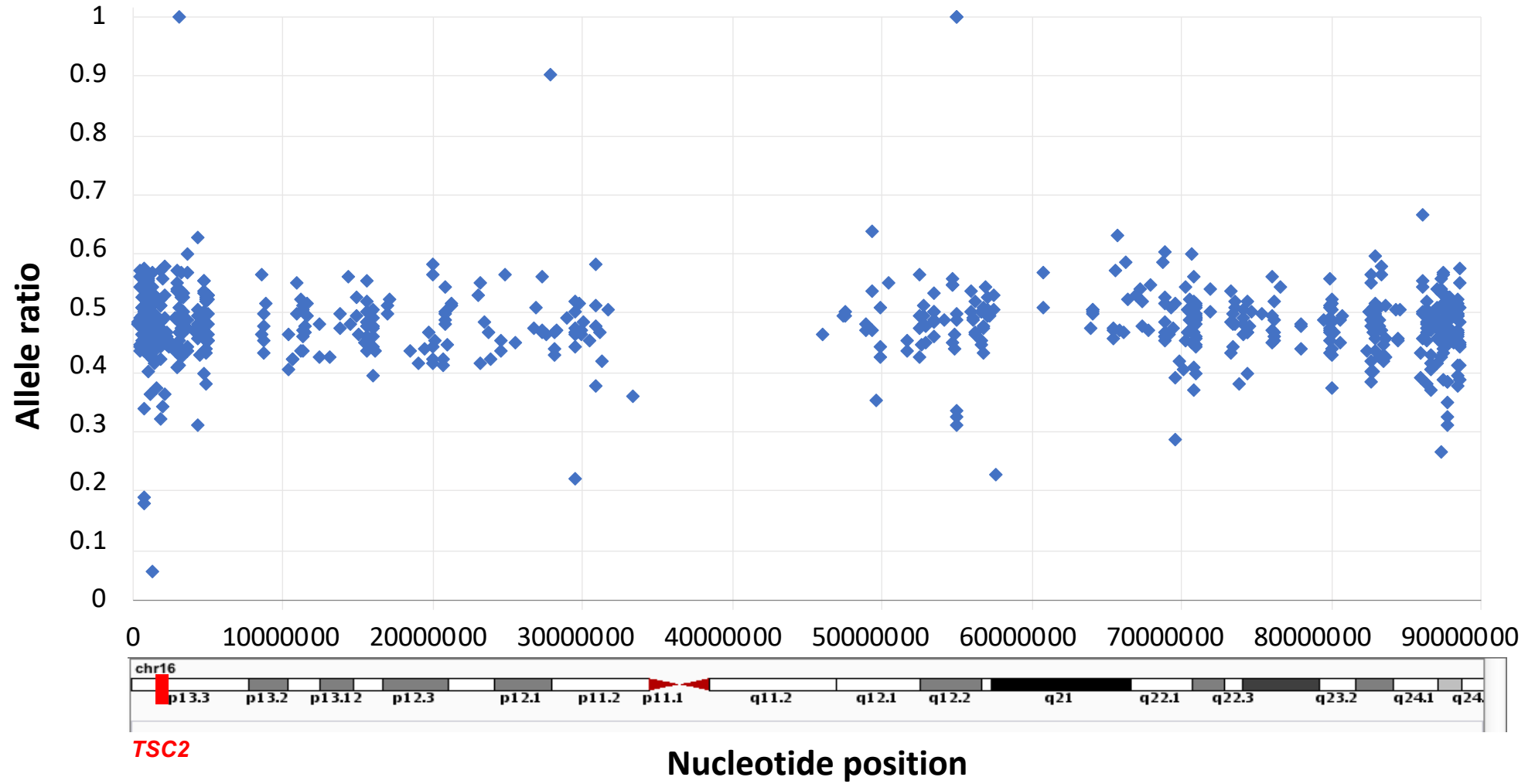
LOH mapping of chromosome 9 in SEGA S9



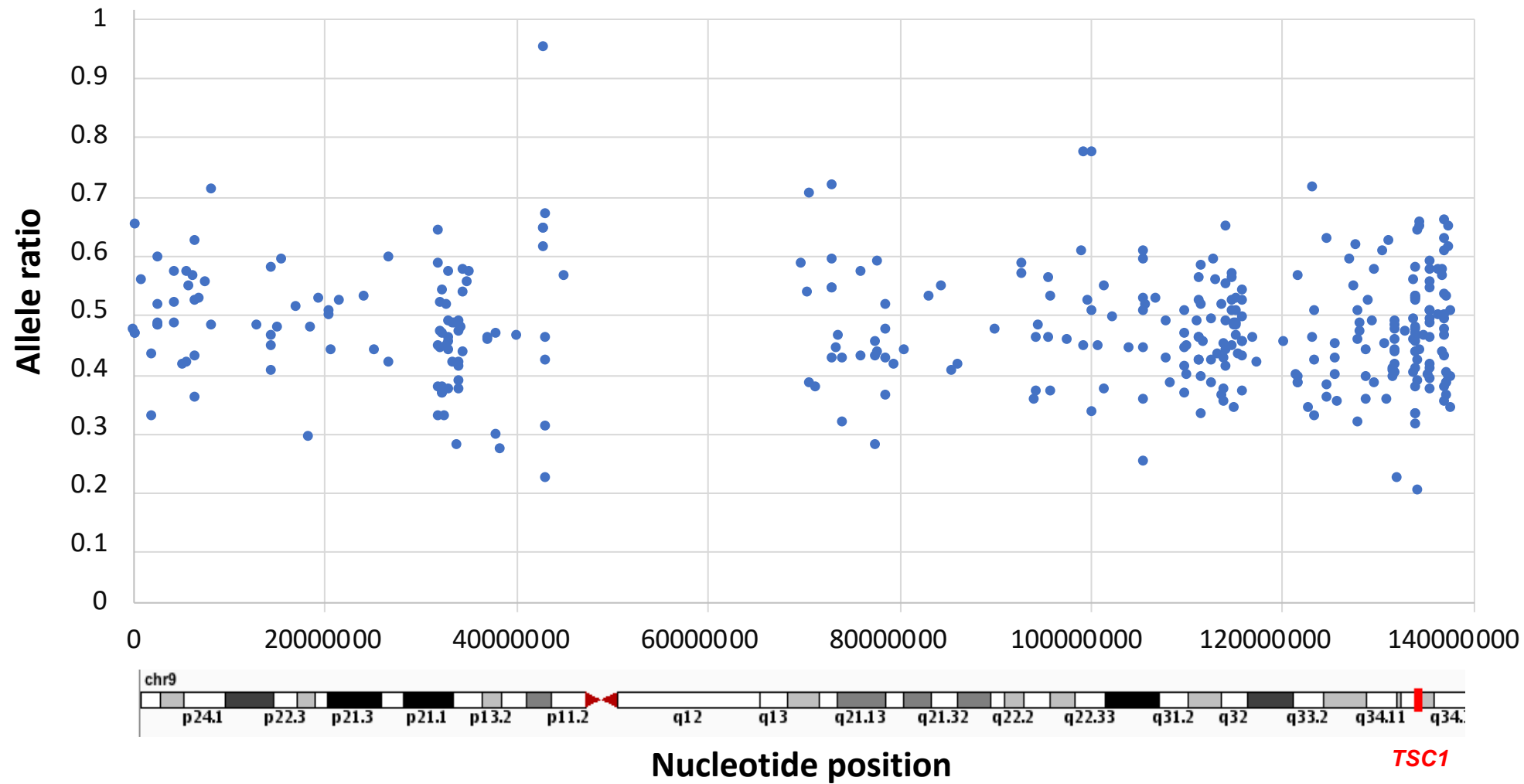
Nucleotide position

TSC1

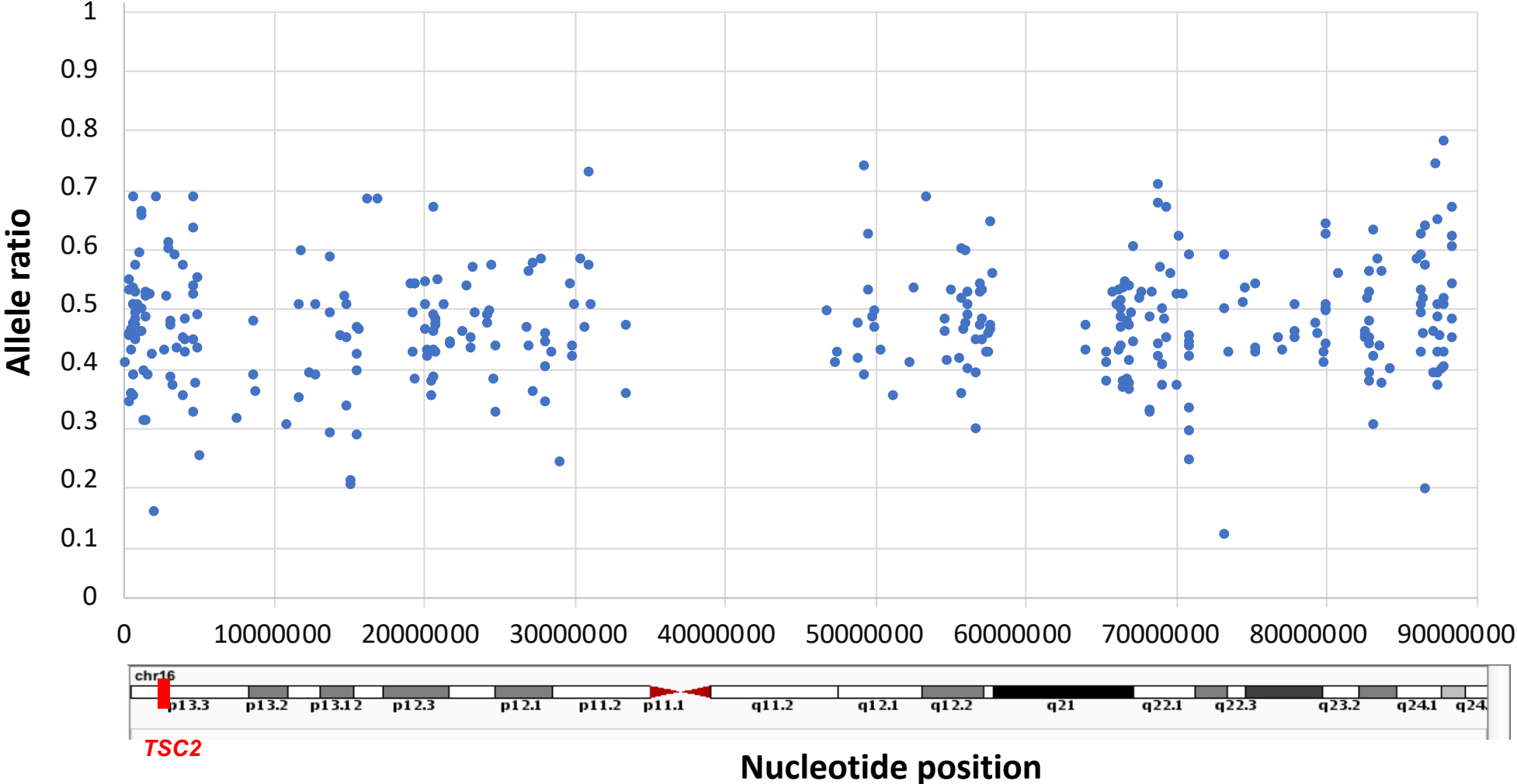
No LOH of chromosome 16 in SEGA S9



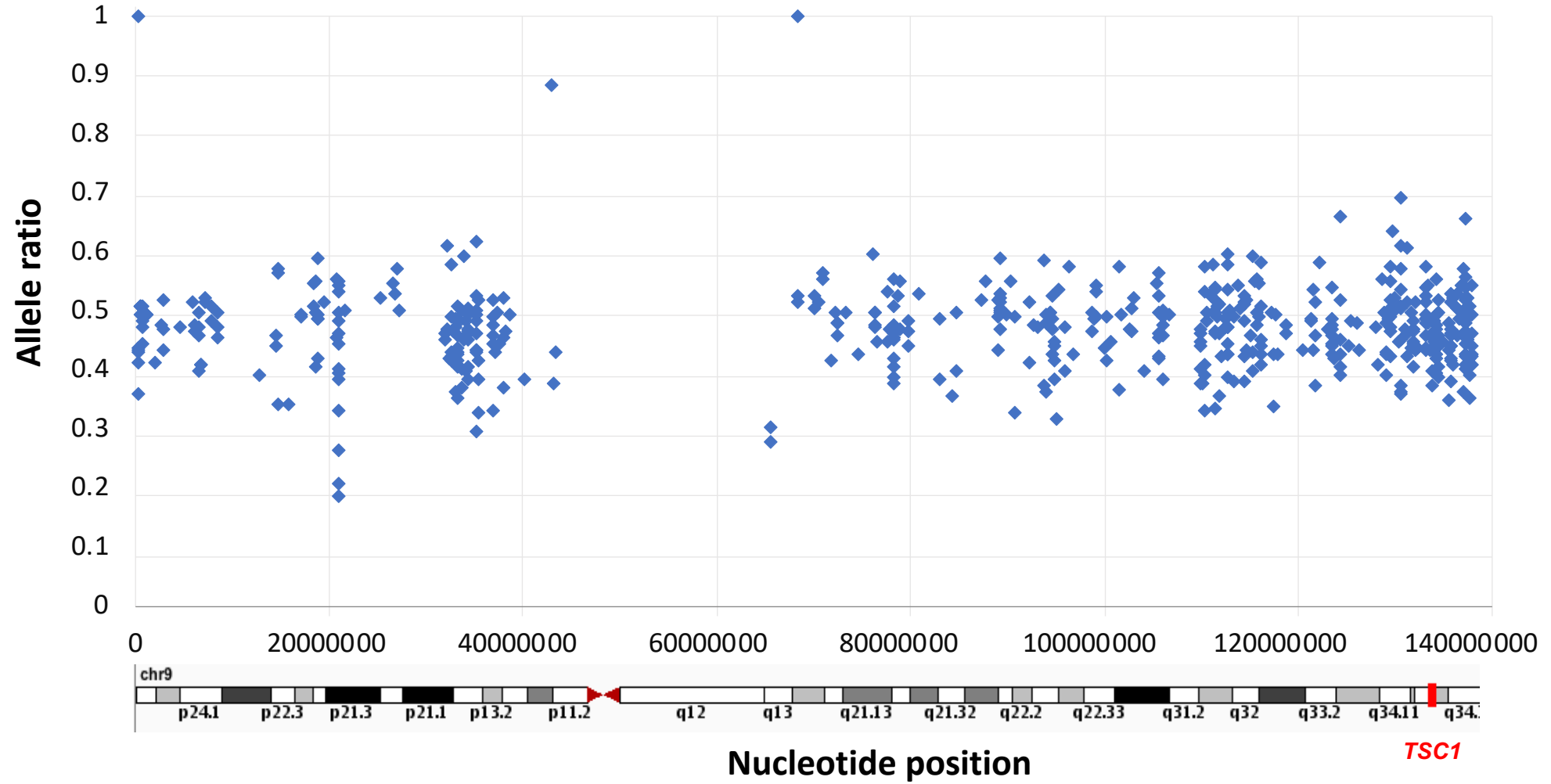
No LOH of chromosome 9 in SEGA S10



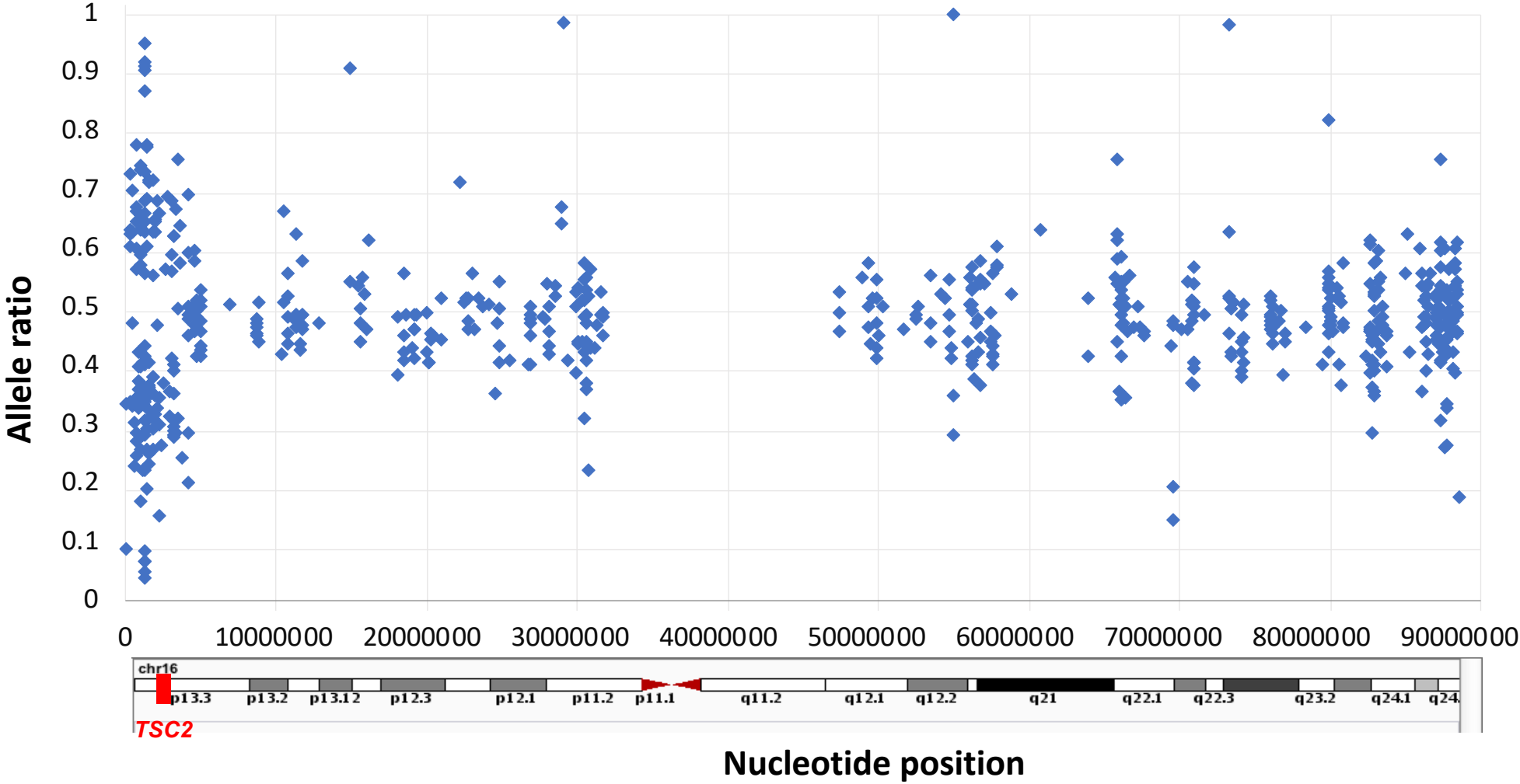
No LOH of chromosome 16 in SEGA S10



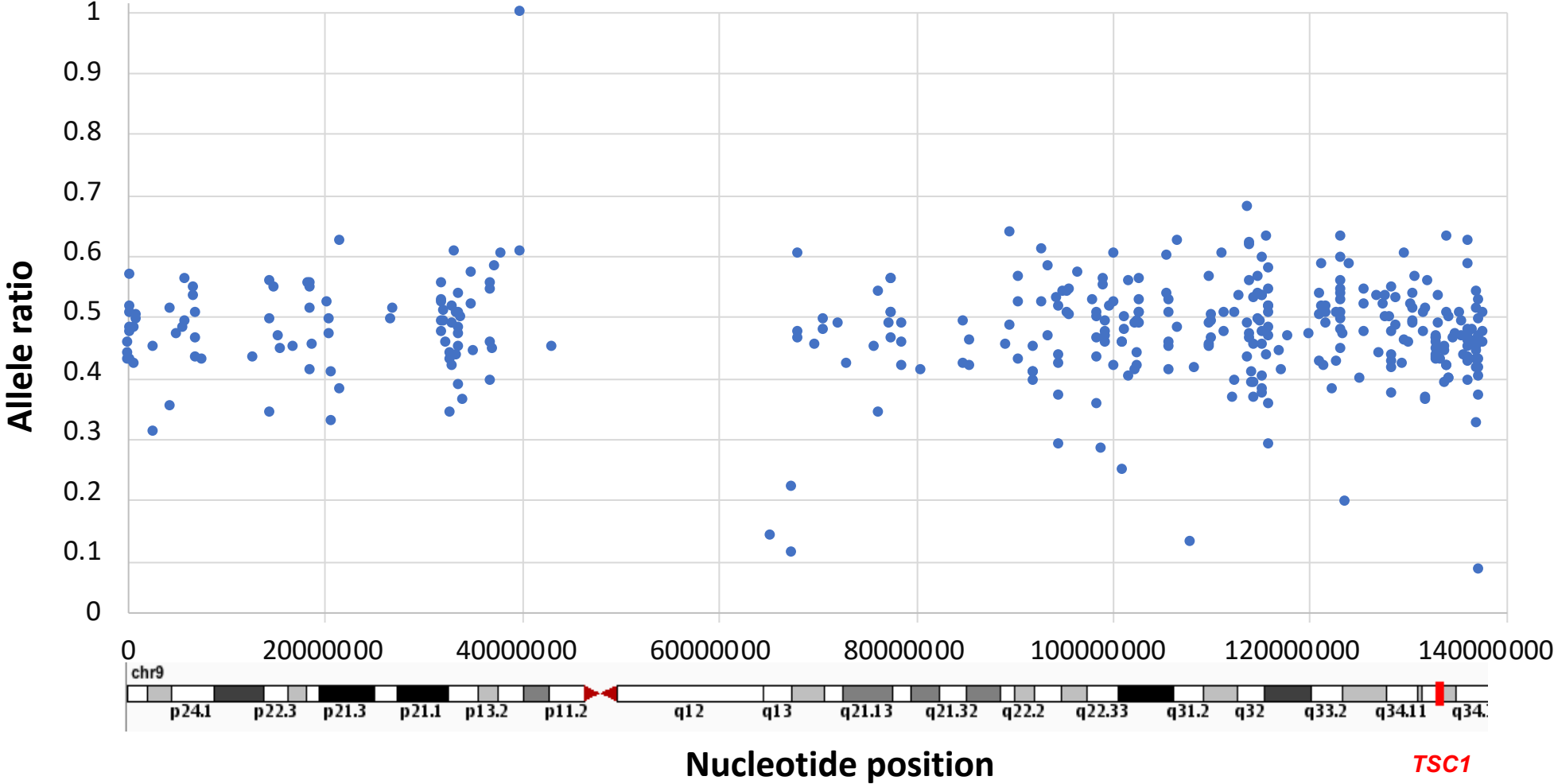
No LOH of chromosome 9 in SEGA S11



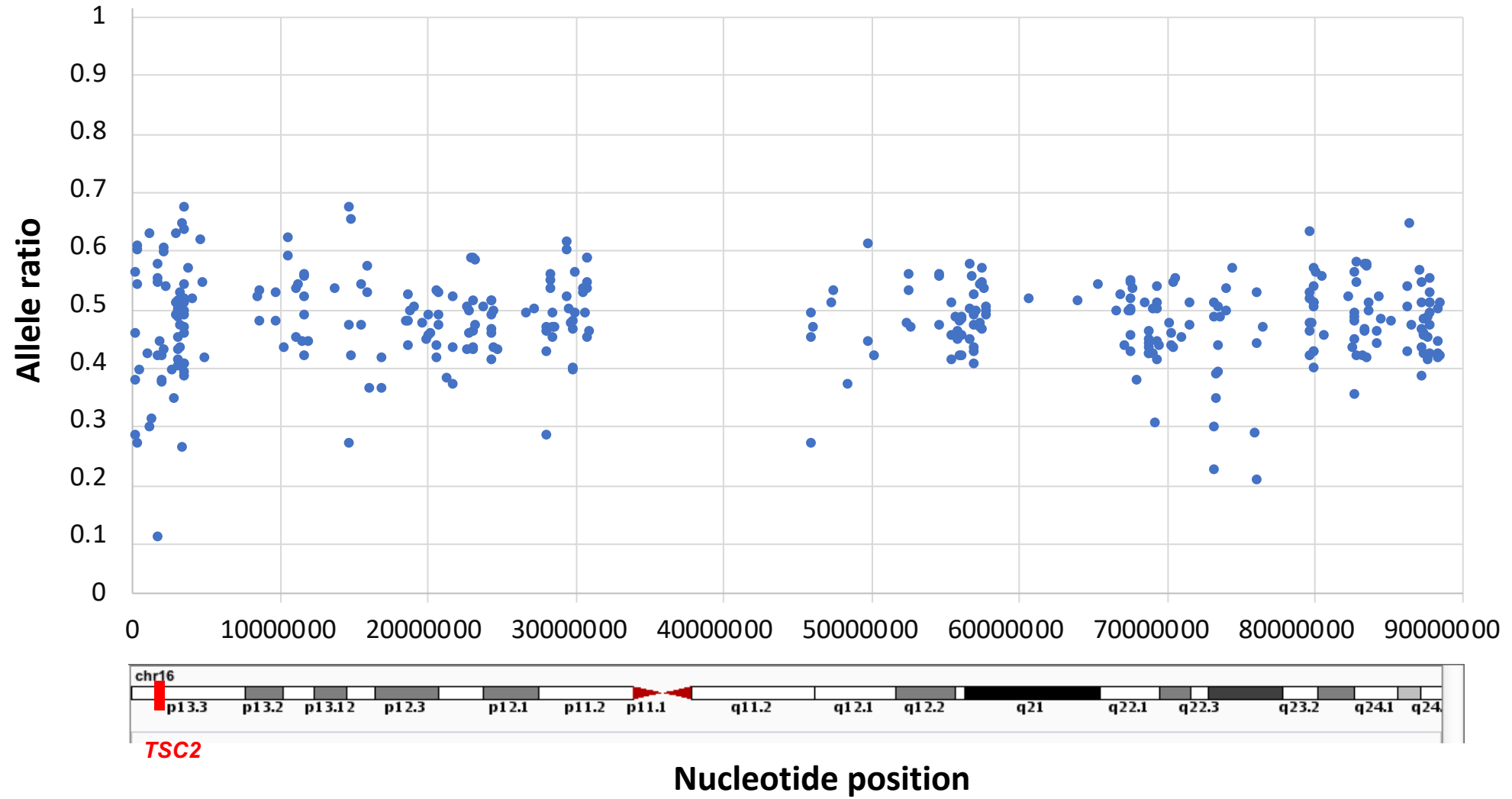
LOH mapping of chromosome 16 in SEGA S11



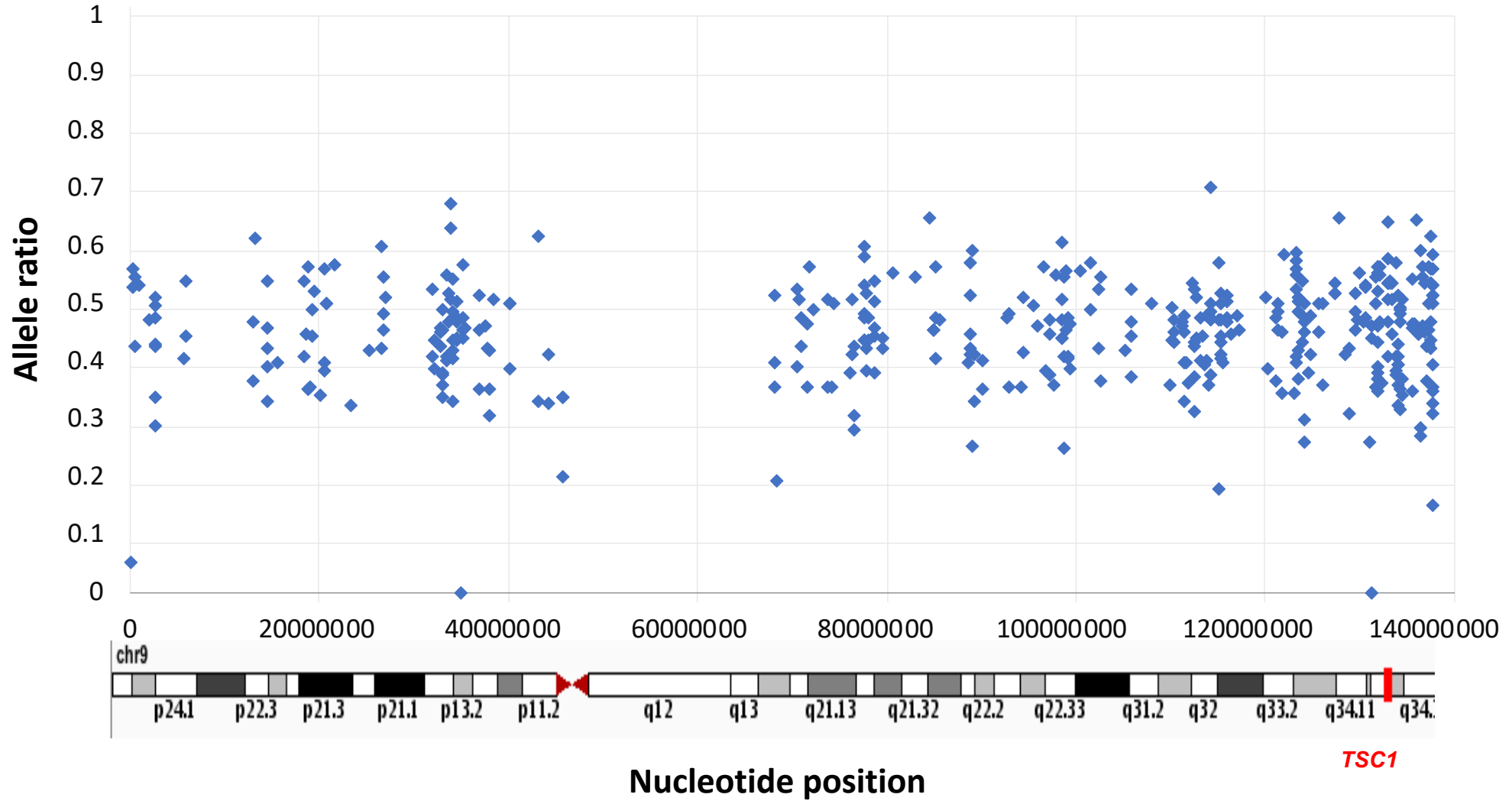
No LOH of chromosome 9 in SEGA S12



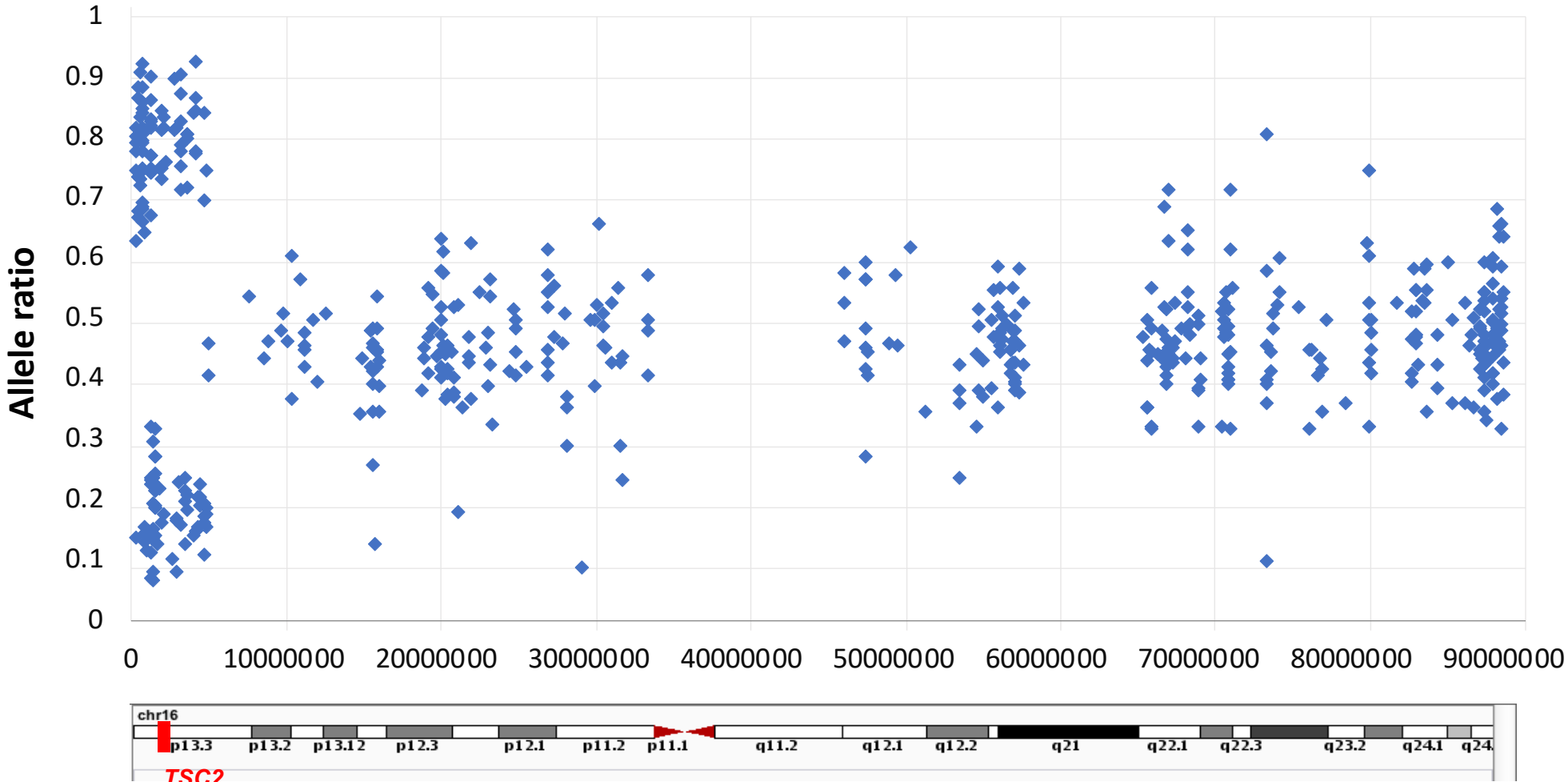
No LOH of chromosome 16 in SEGA S12



No LOH of chromosome 9 in SEGA S13

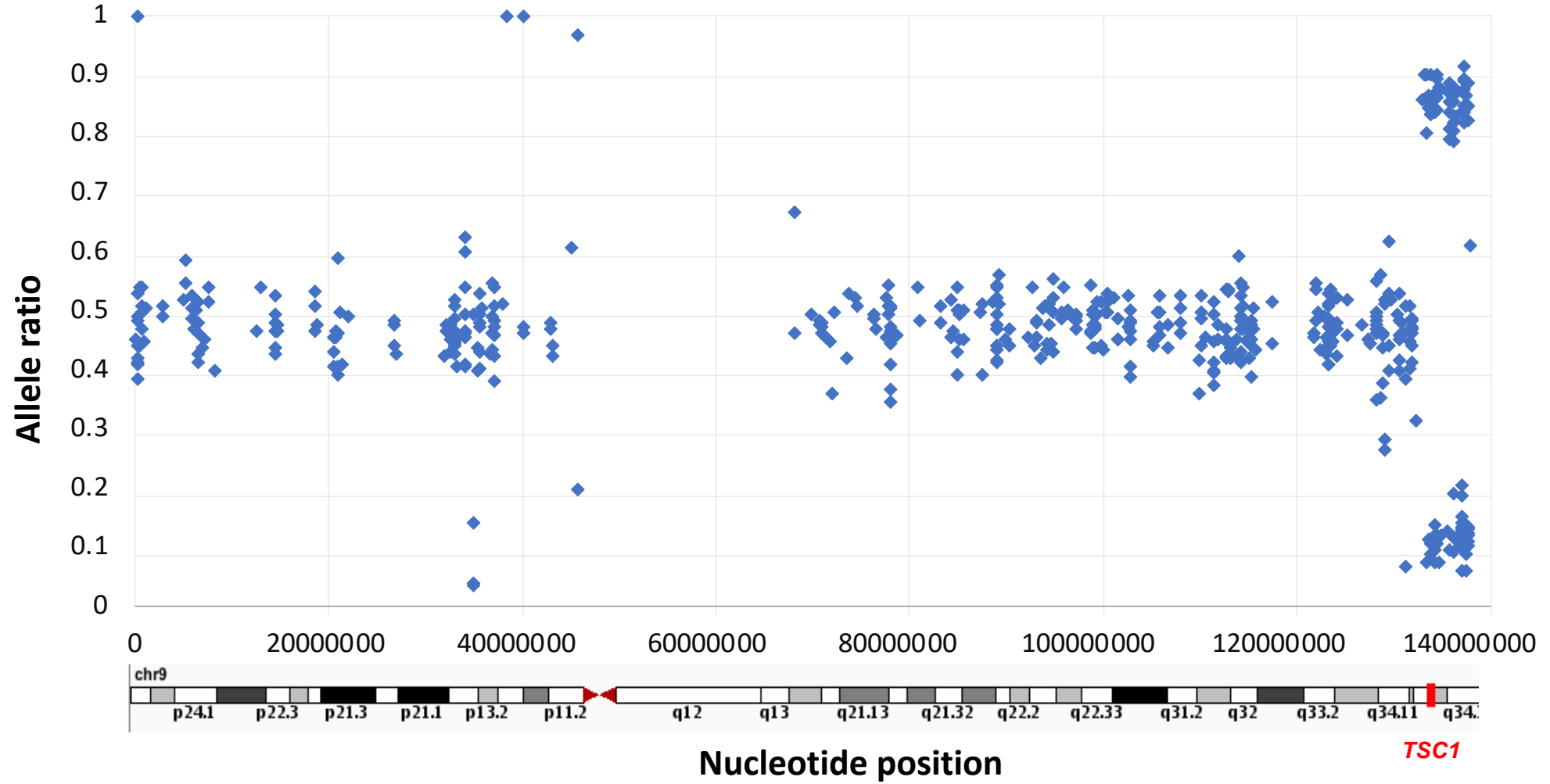


LOH mapping of chromosome 16 in SEGA S13

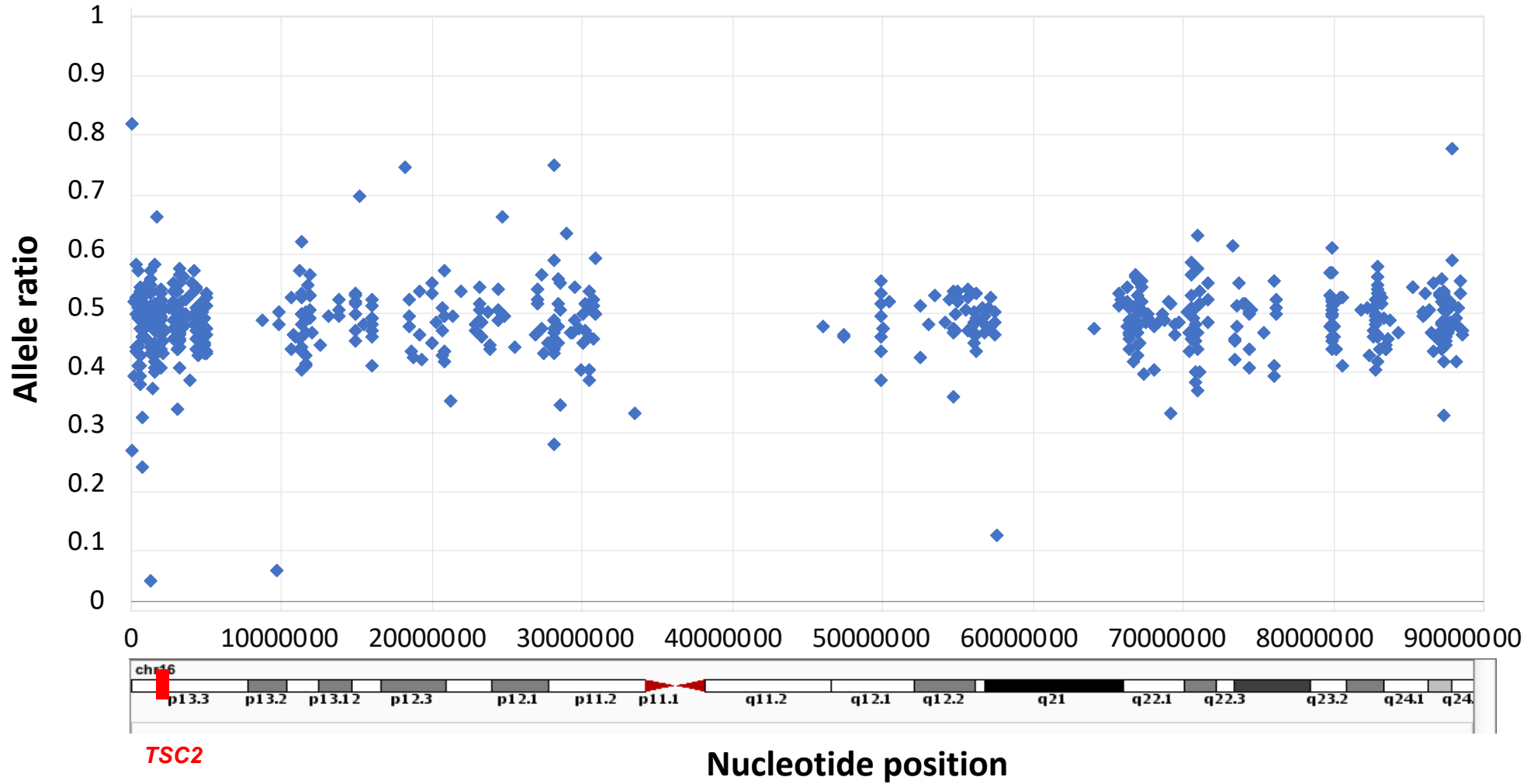


Nucleotide position

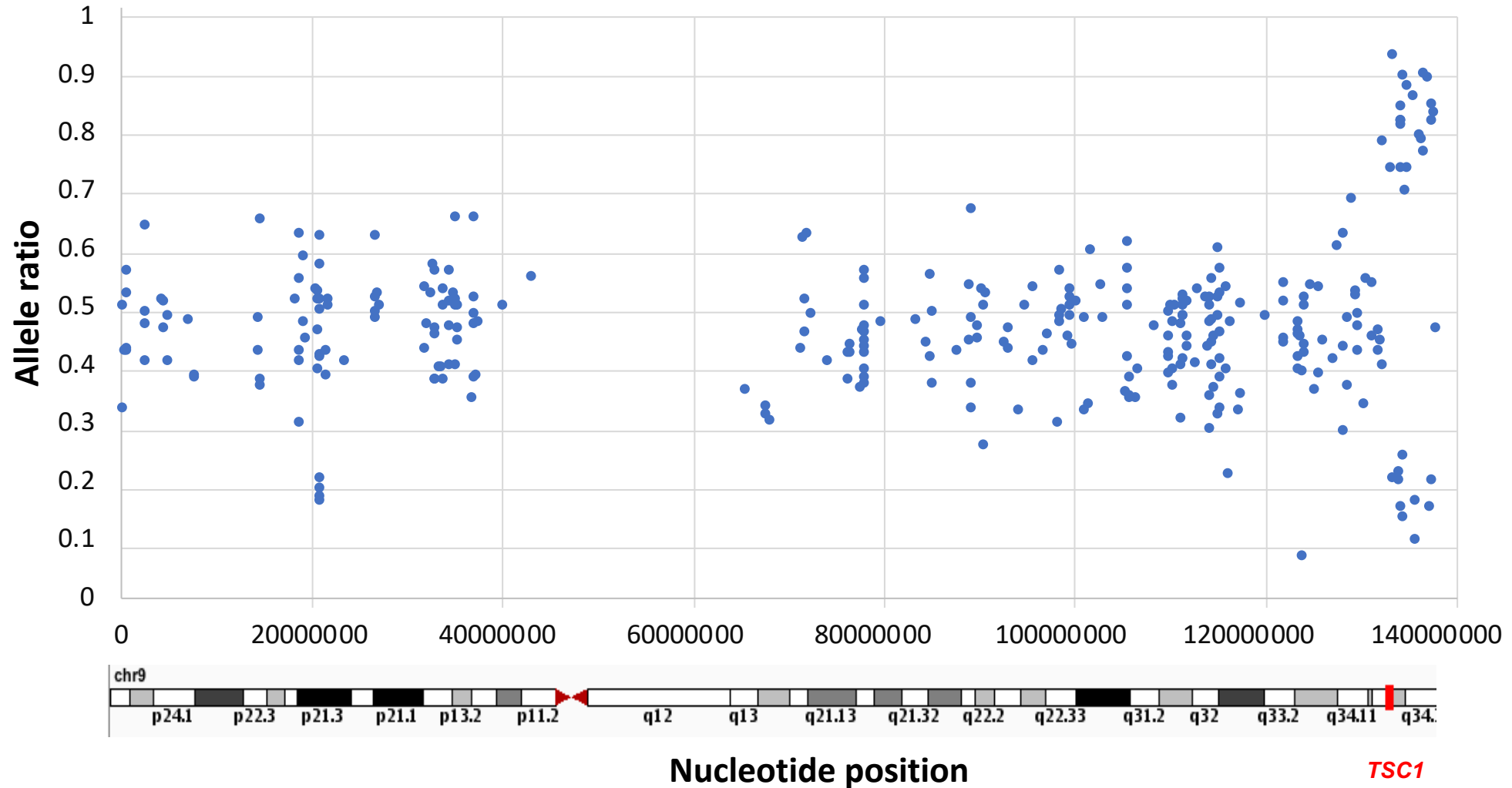
LOH mapping of chromosome 9 in SEGA S14



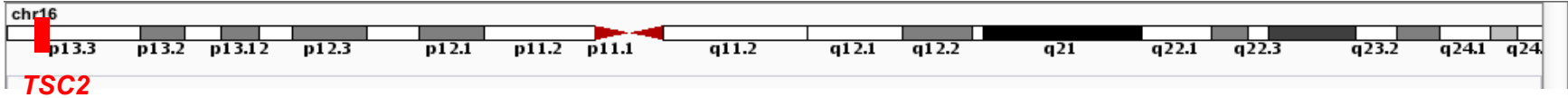
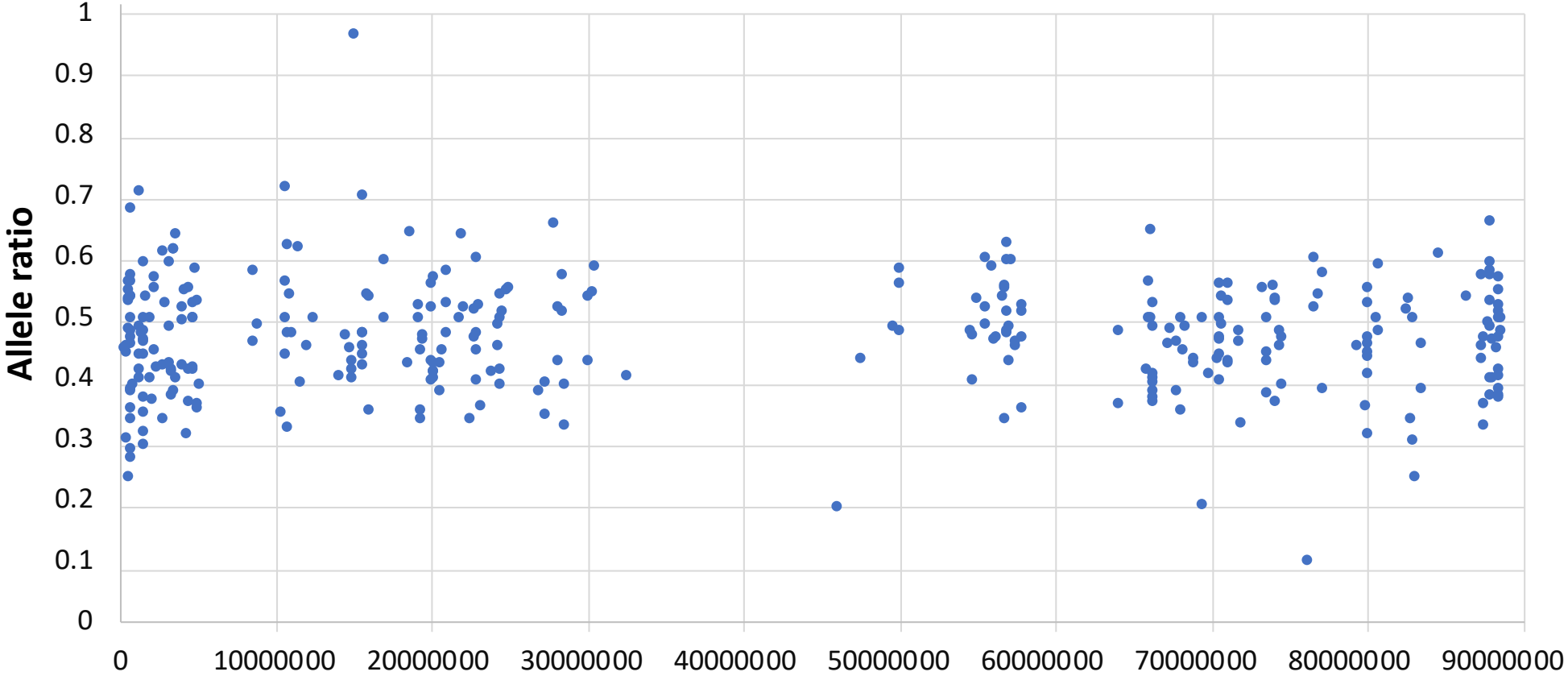
No LOH of chromosome 16 in SEGA S14



LOH mapping of chromosome 9 in SEGA S15

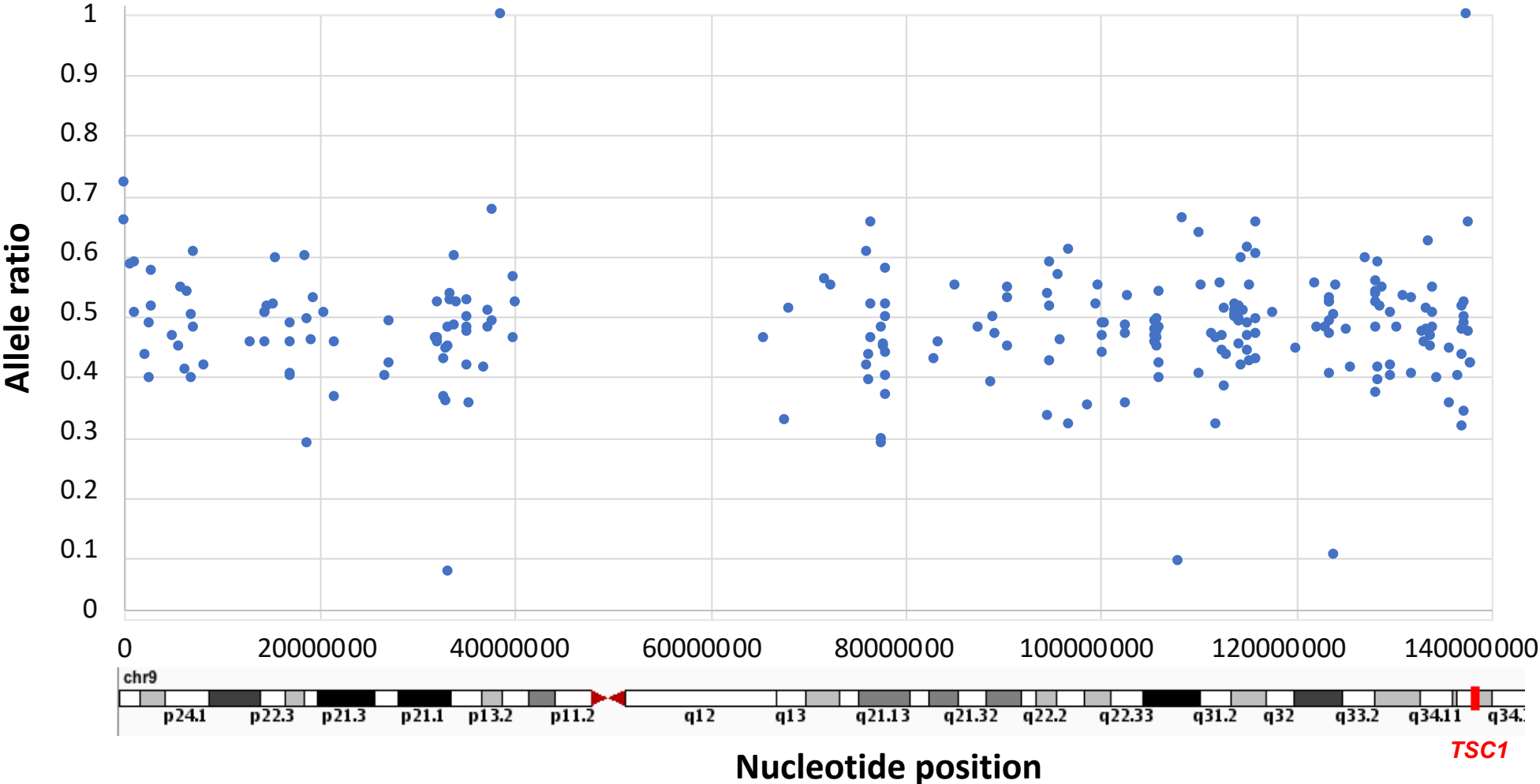


No LOH of chromosome 16 in SEGA S15

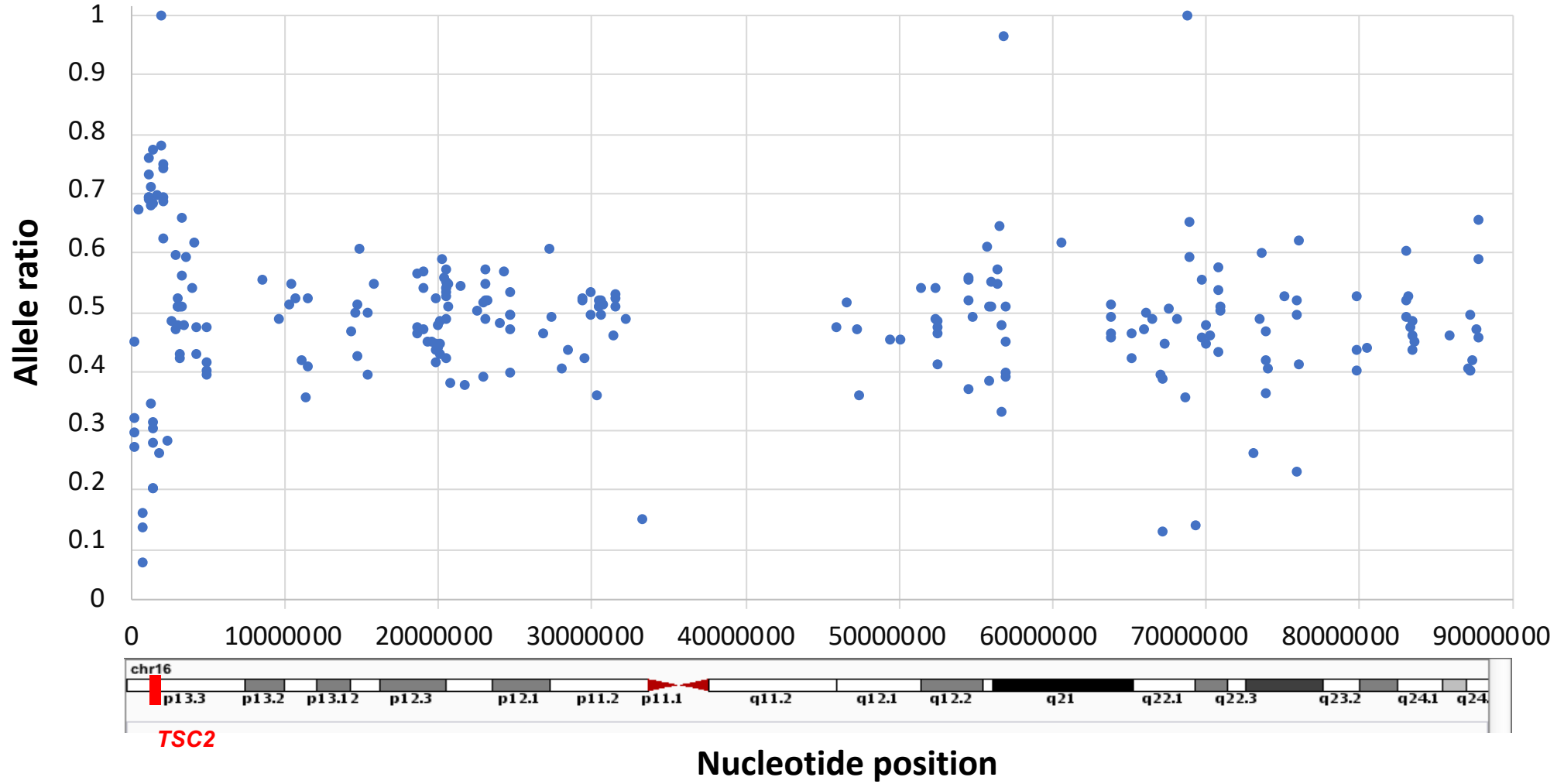


Nucleotide position

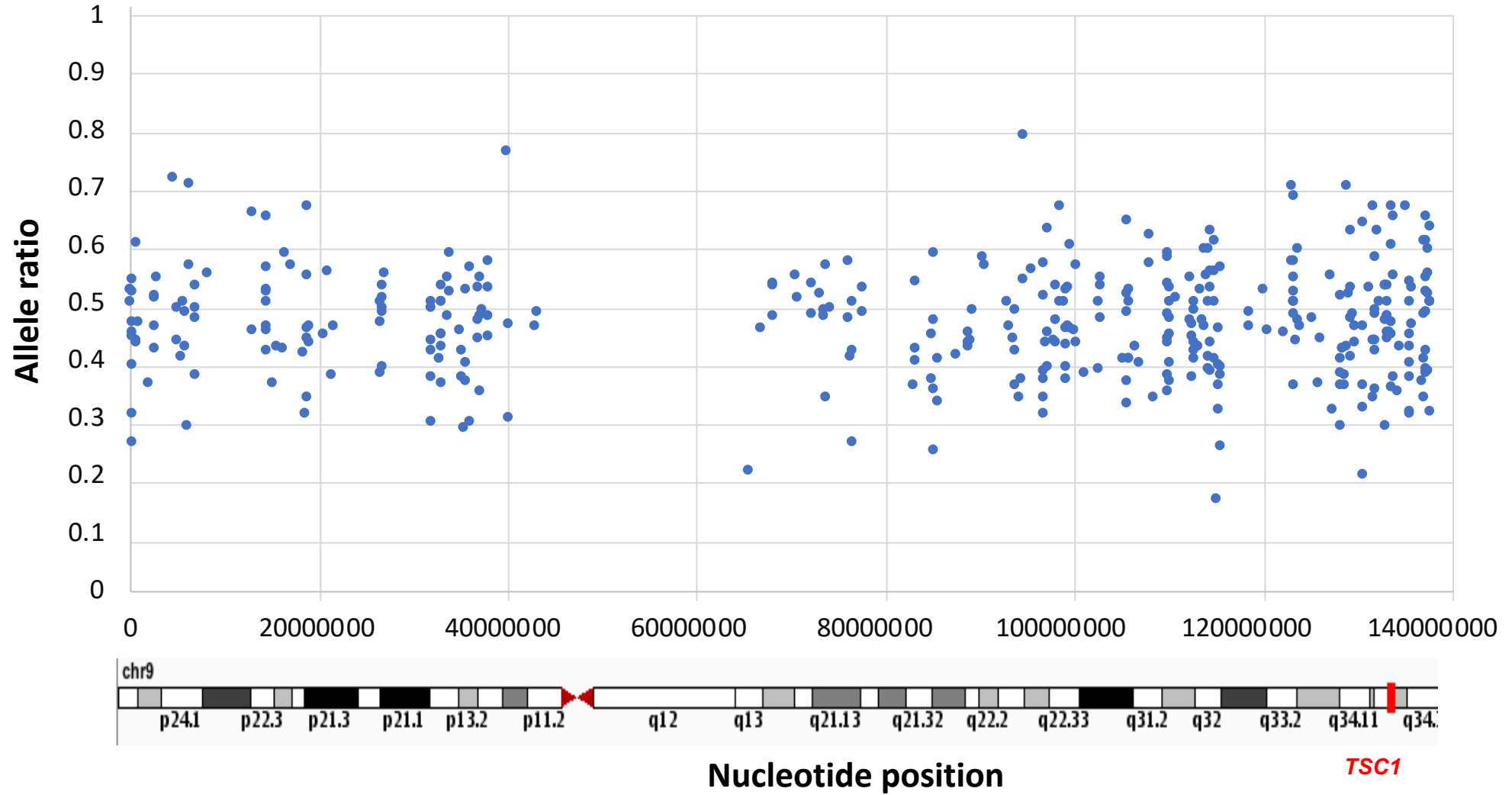
No LOH of chromosome 9 in SEGA S16



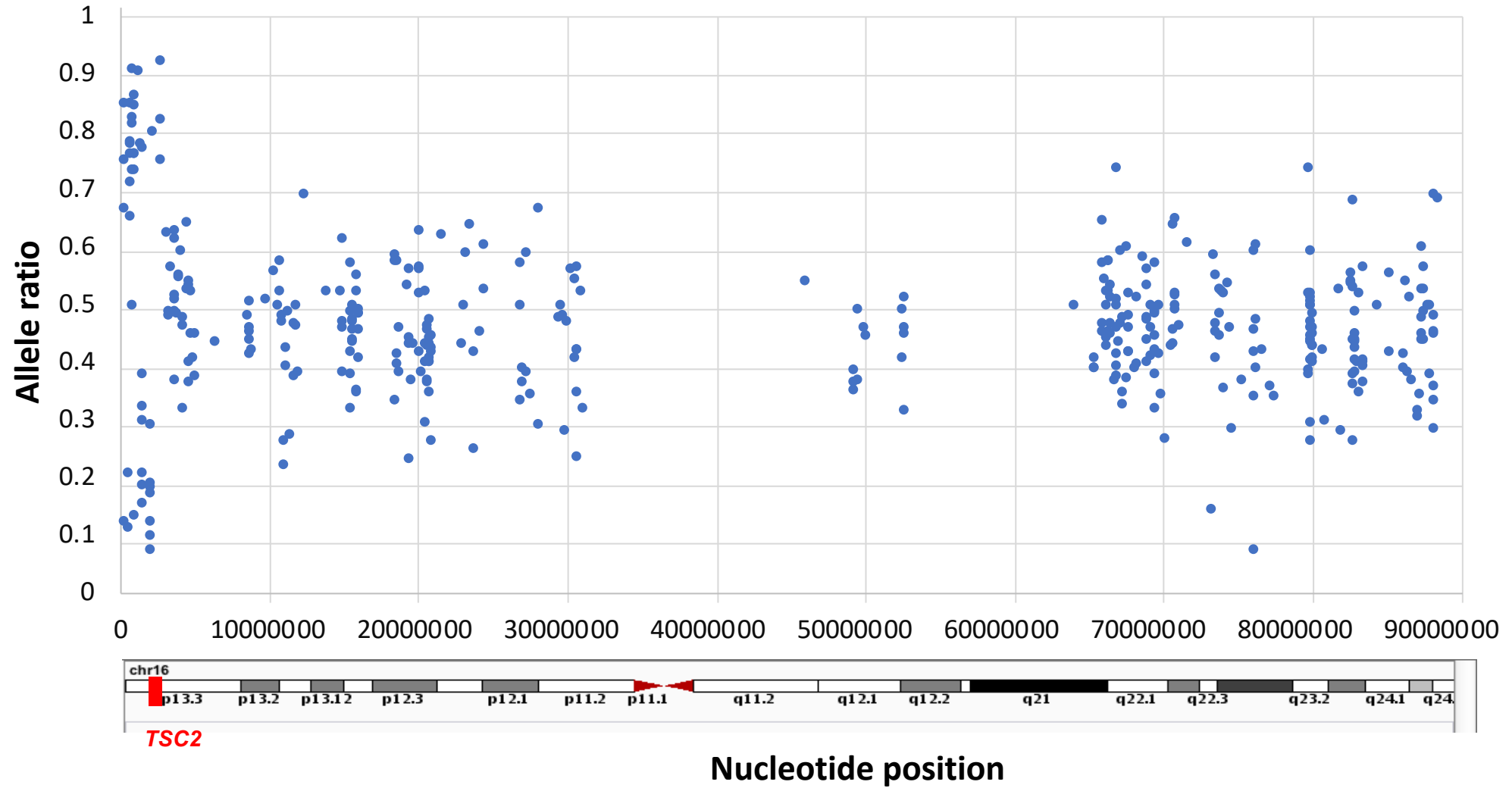
LOH mapping of chromosome 16 in SEGA S16



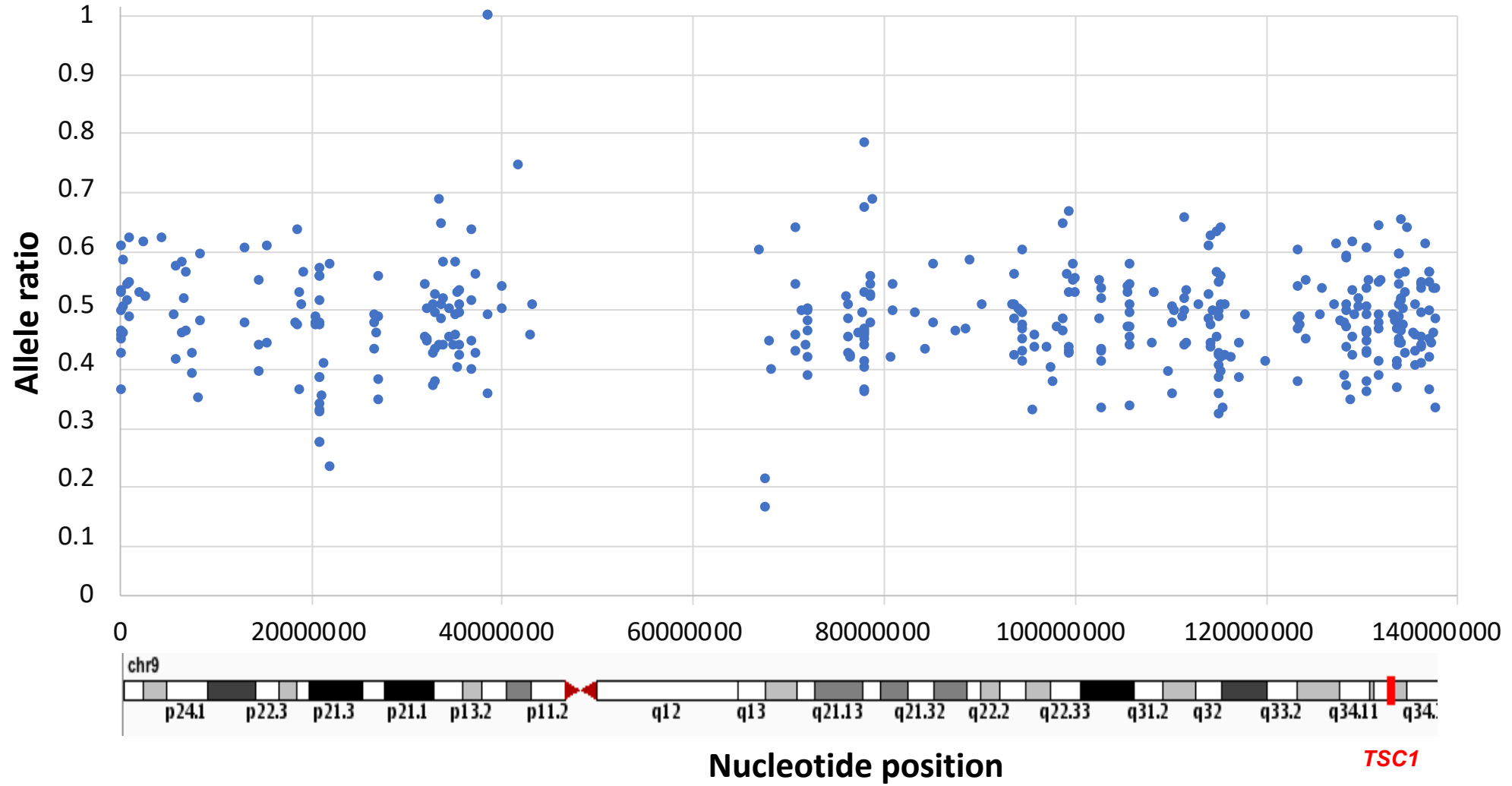
No LOH of chromosome 9 in SEGA S17



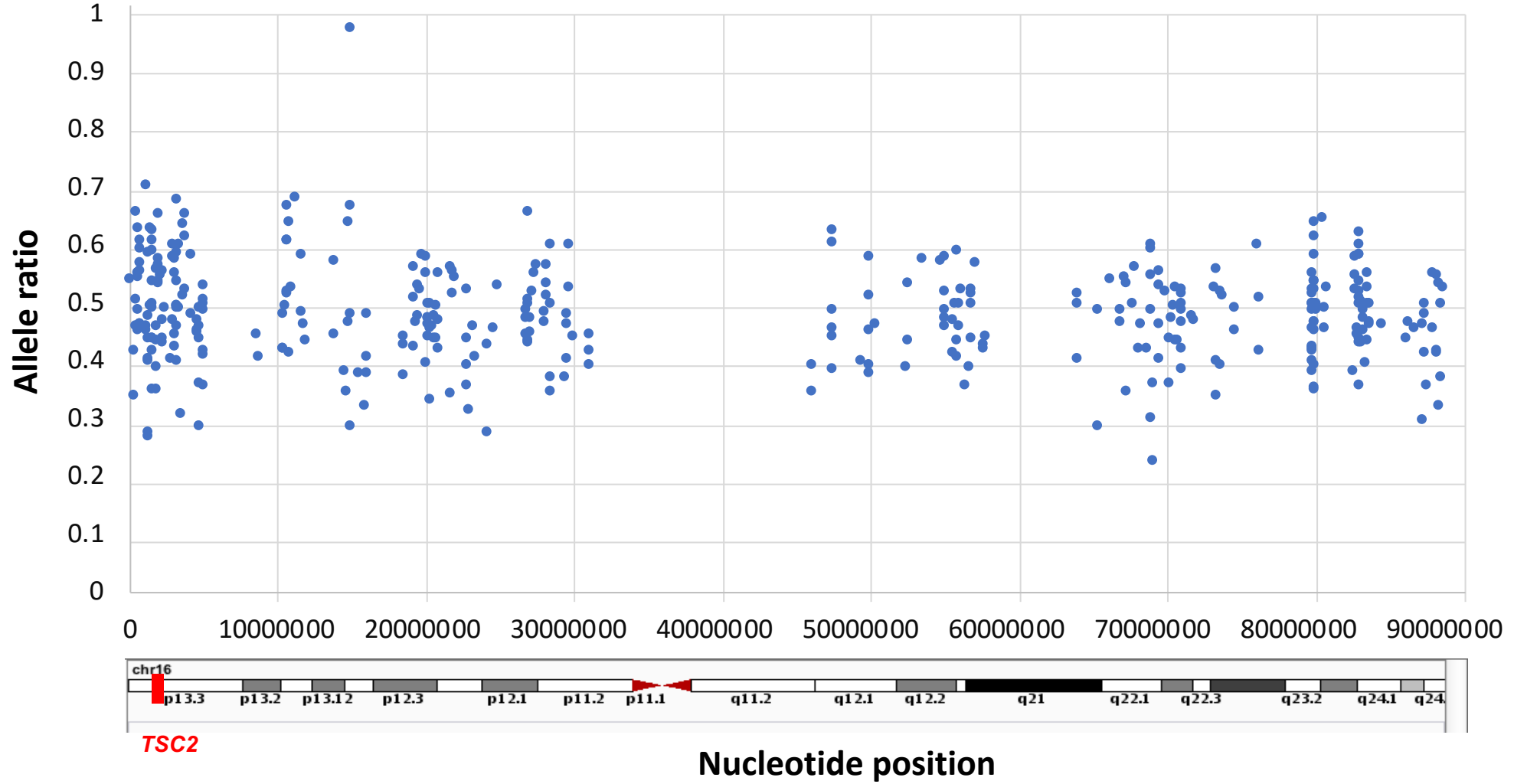
LOH mapping of chromosome 16 in SEGA S17



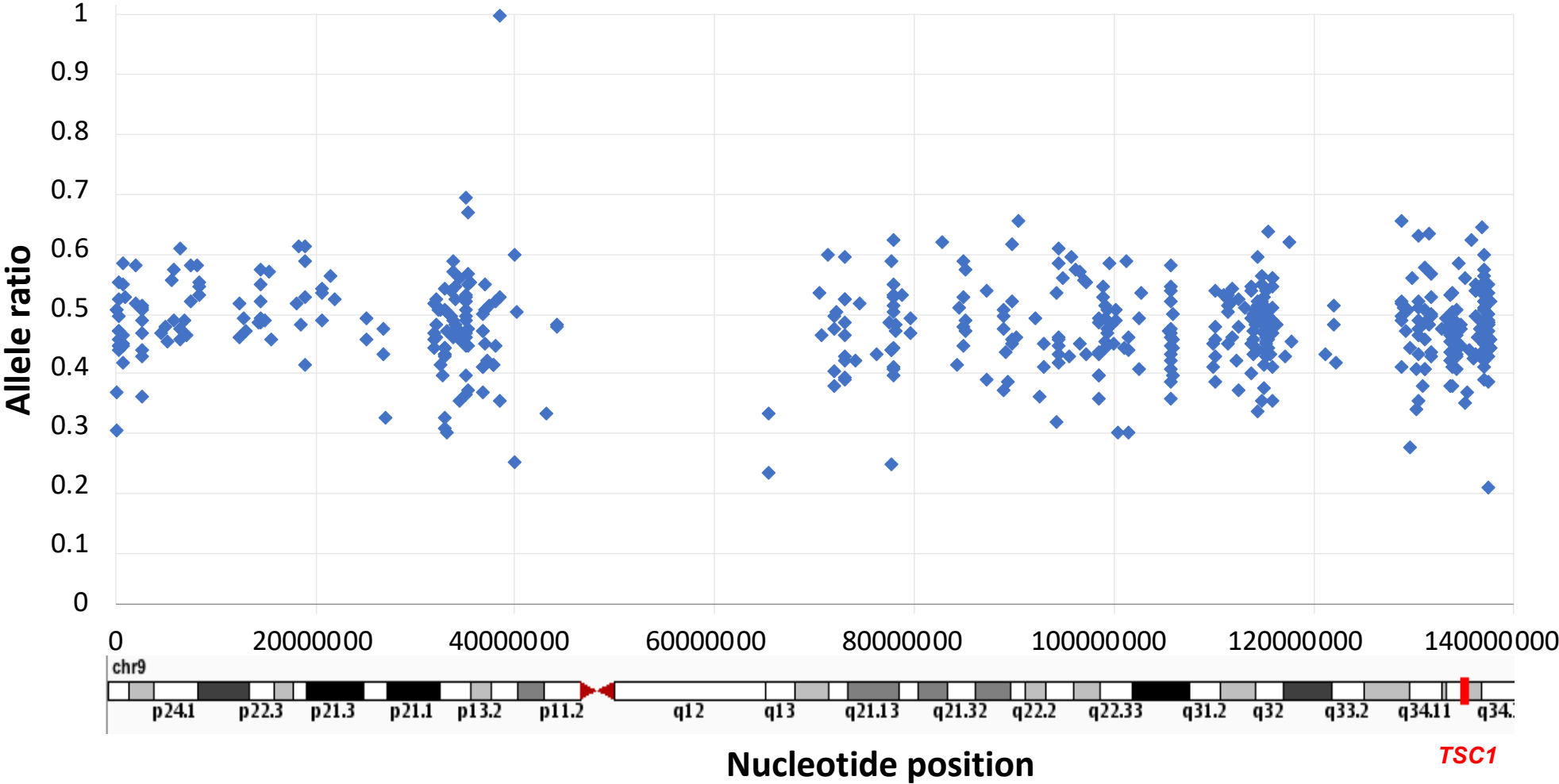
No LOH of chromosome 9 in SEGA S18



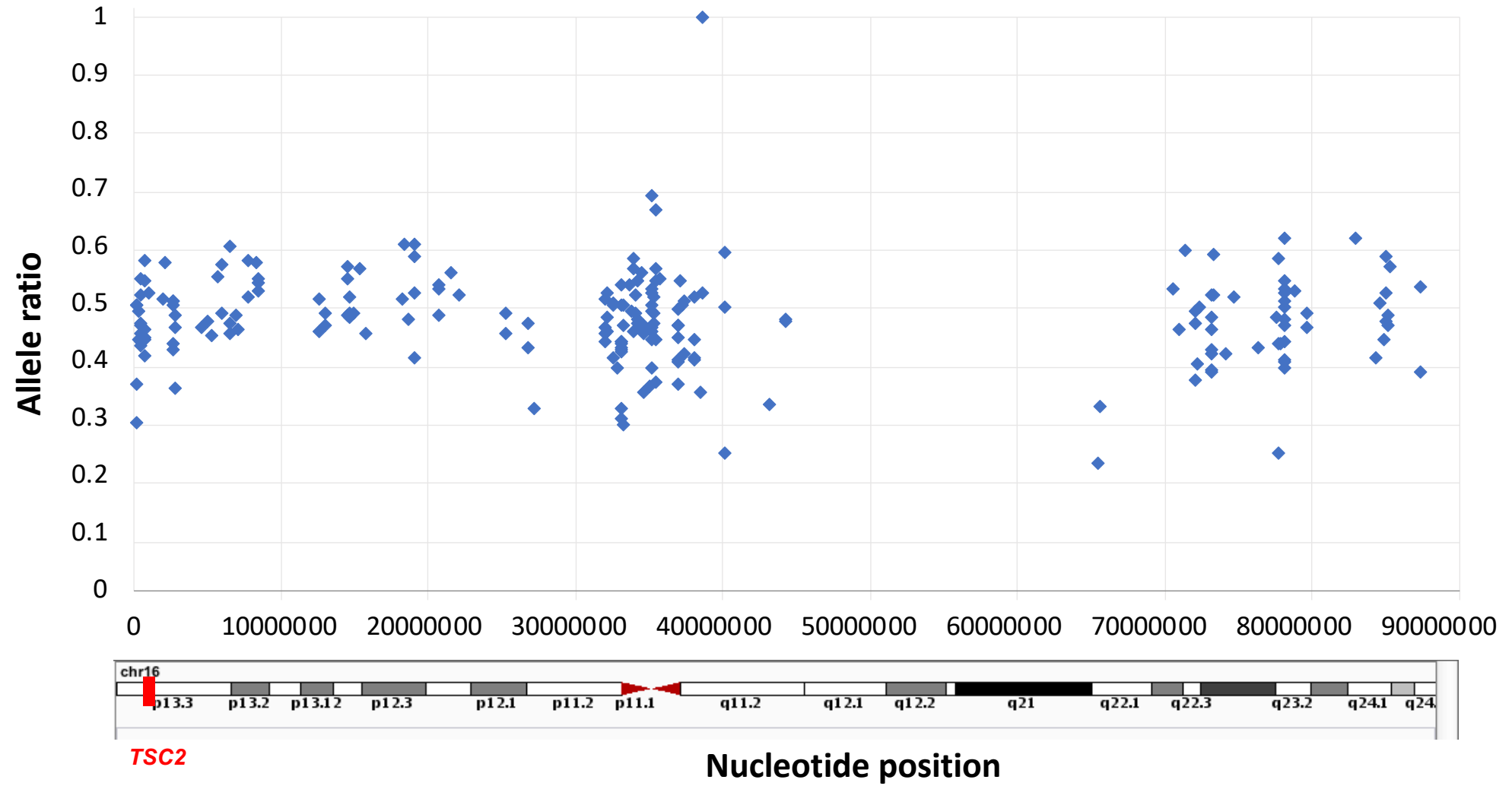
No LOH of chromosome 16 in SEGA S18



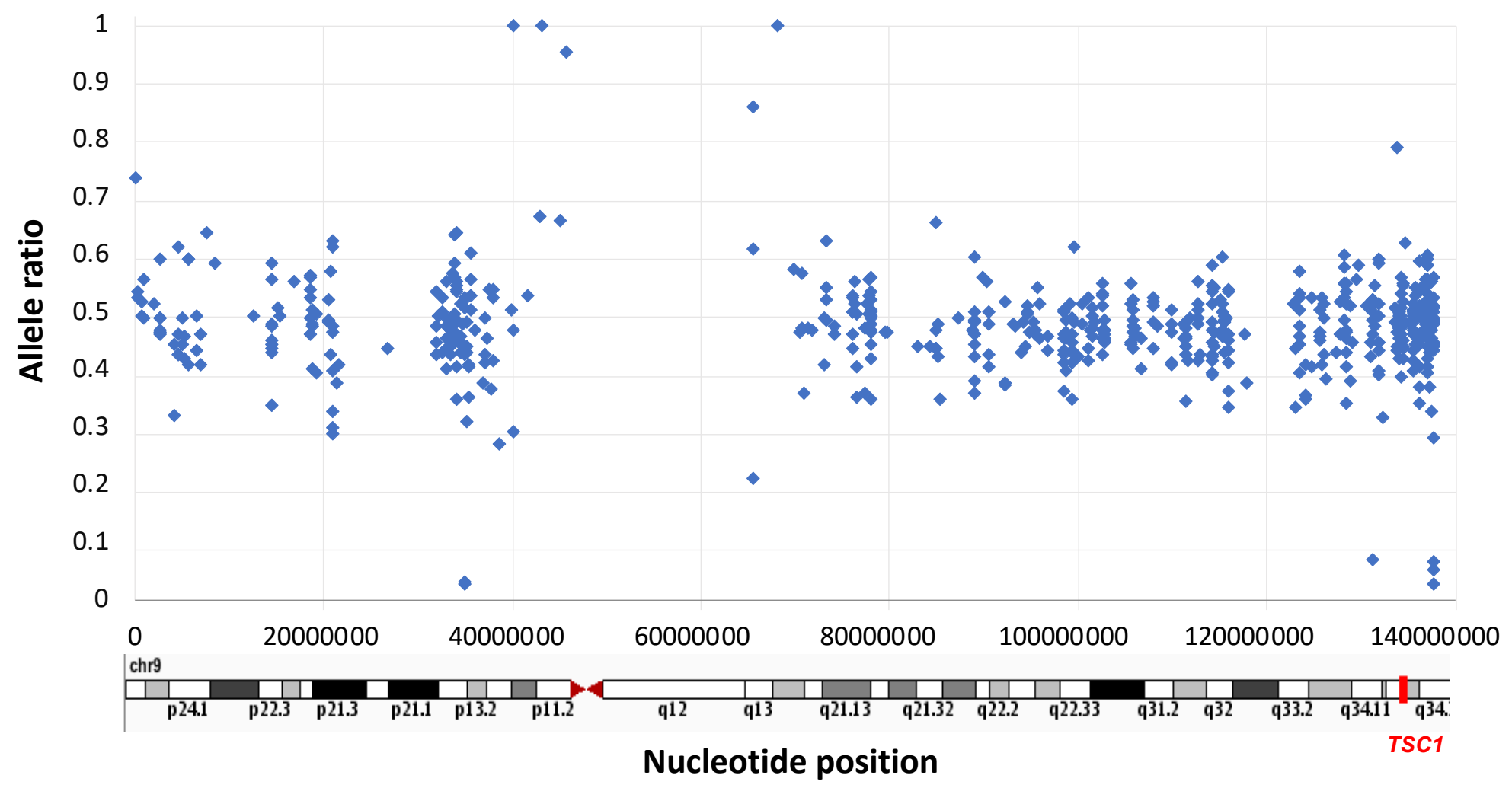
No LOH of chromosome 9 in SEGA S19



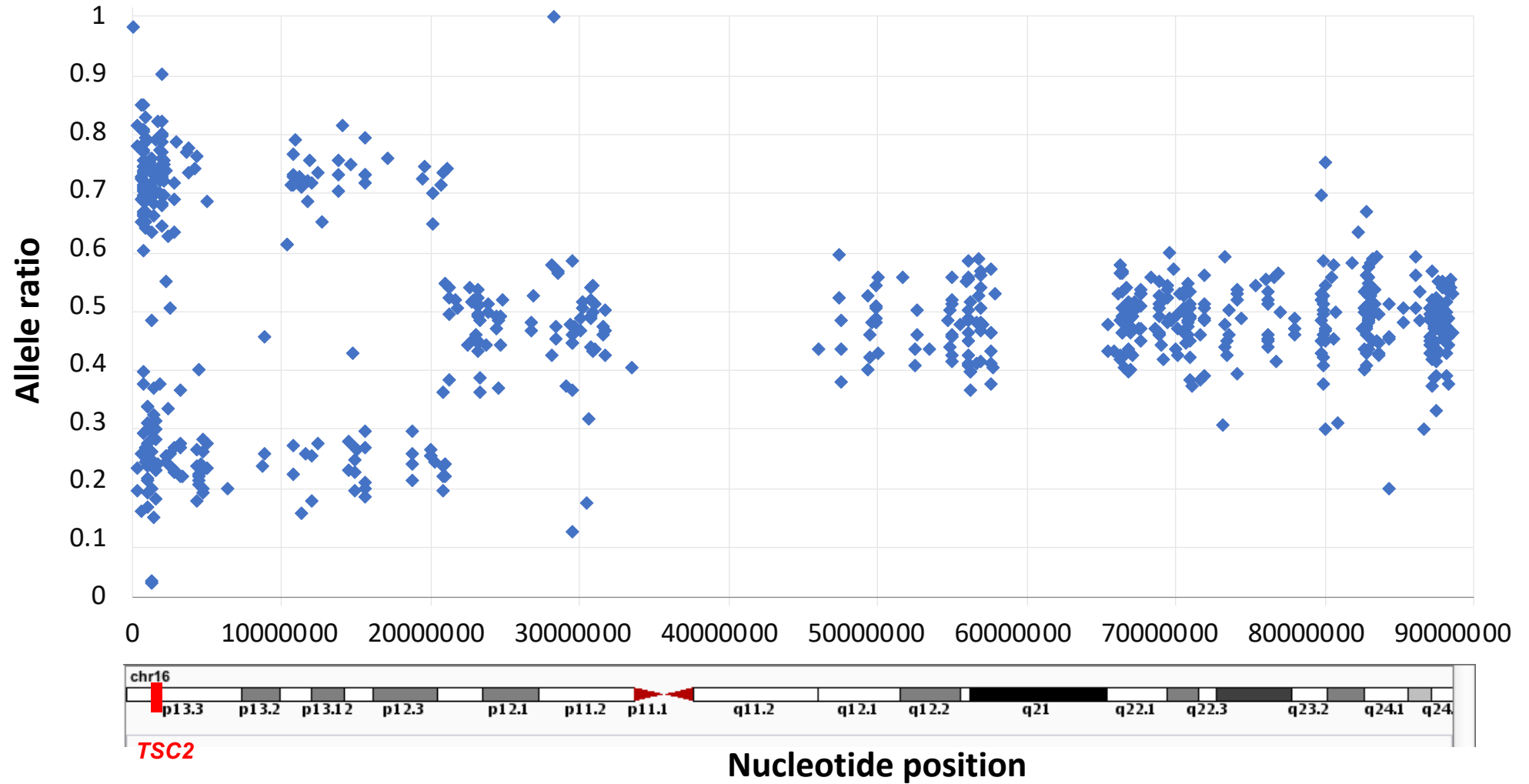
No LOH of chromosome 16 in SEGA S19



No LOH of chromosome 9 in SEGA S20



LOH mapping of chromosome 16 in SEGA S20



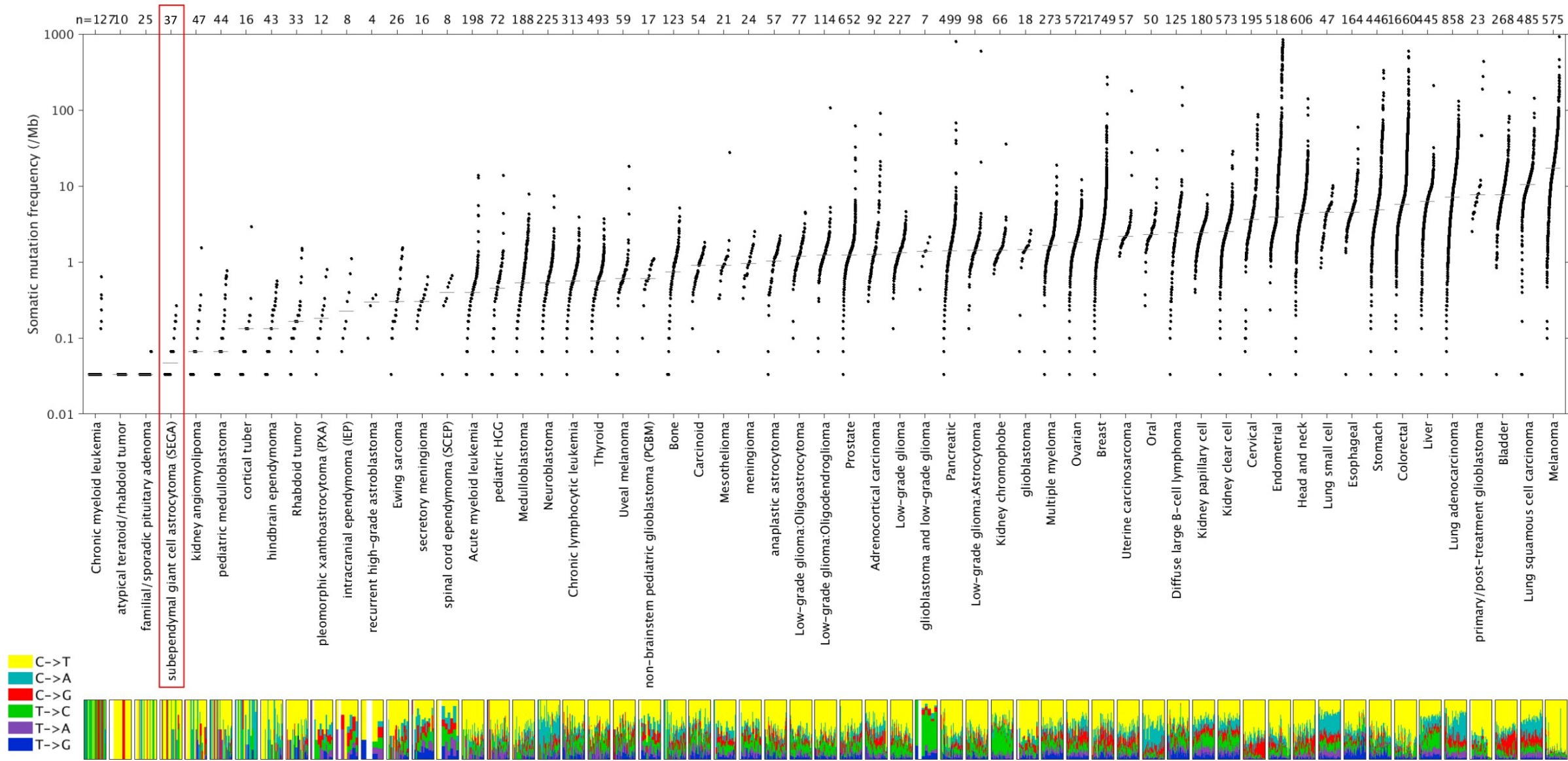
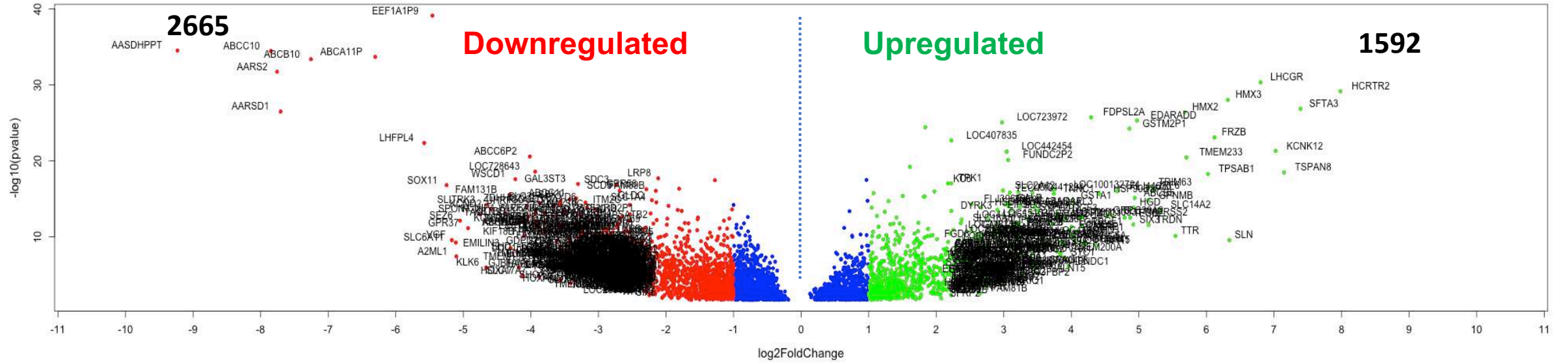
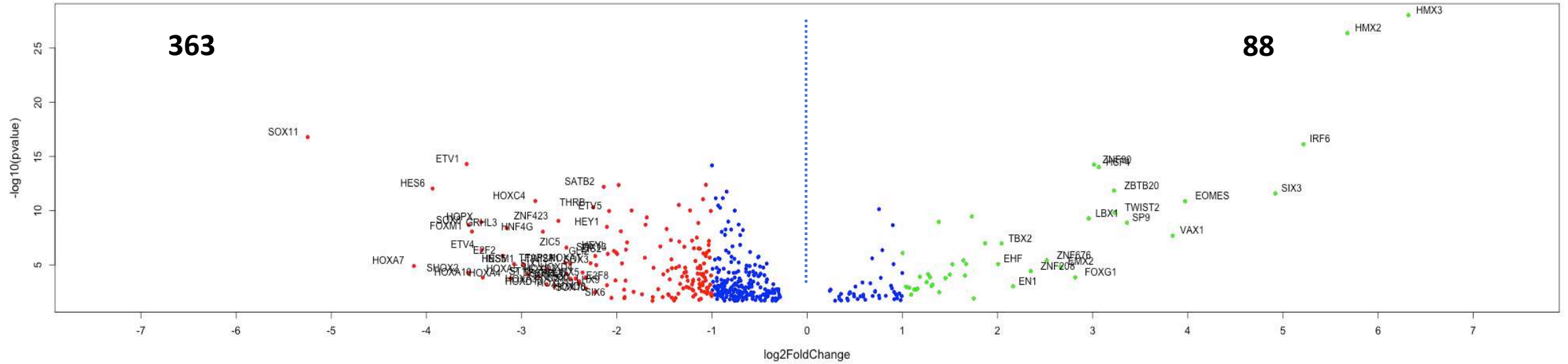


Fig. S2

Distribution of all differentially expressed genes in SEGA and TCGA astrocytomas

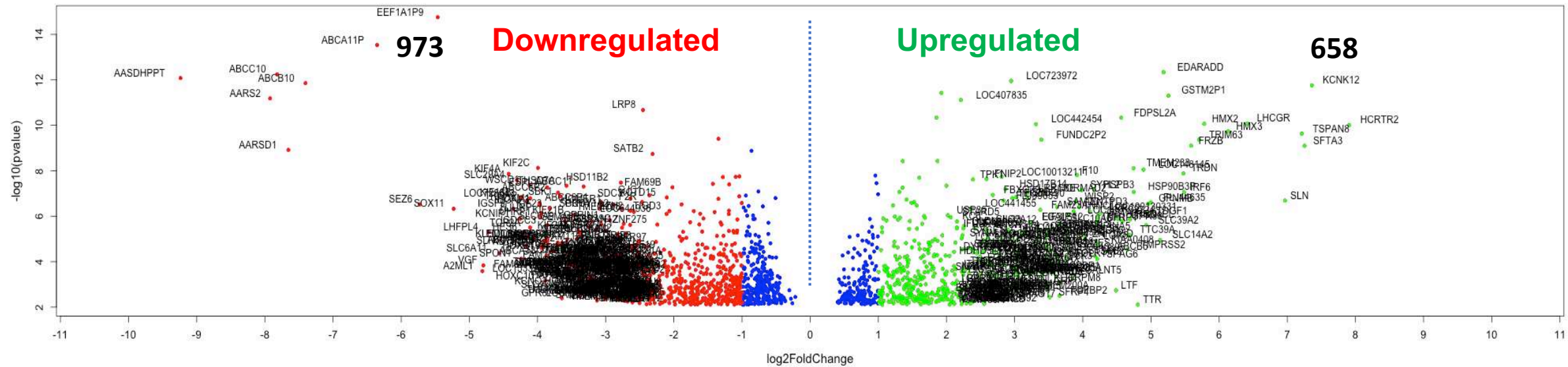


Distribution of all differentially expressed TFs in SEGA and TCGA astrocytomas



FDR < 0.05

Distribution of all differentially expressed genes in SEGA and TCGA oligodendrogliomas



Distribution of all differentially expressed TFs in SEGA and TCGA oligodendrogliomas

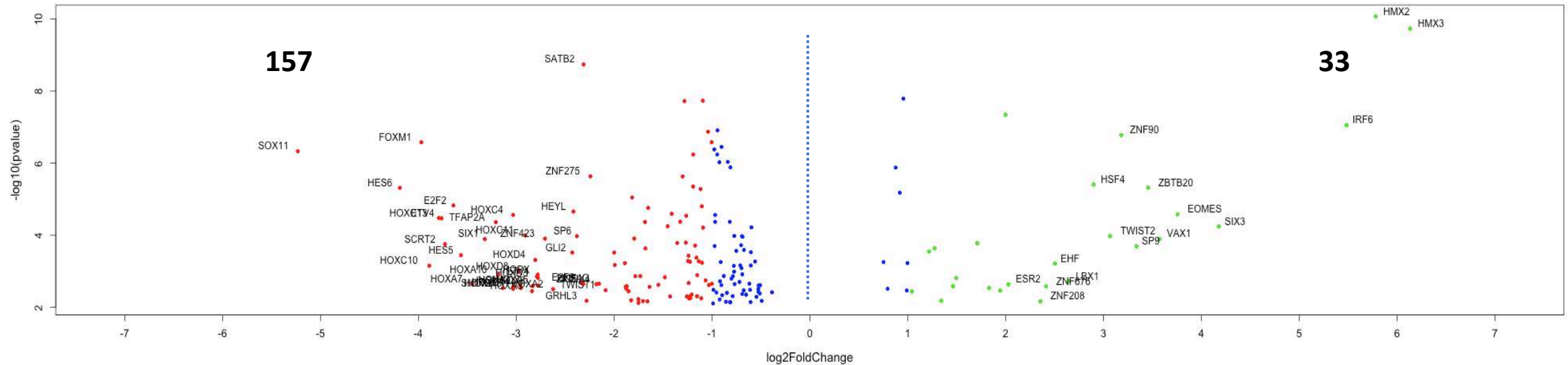
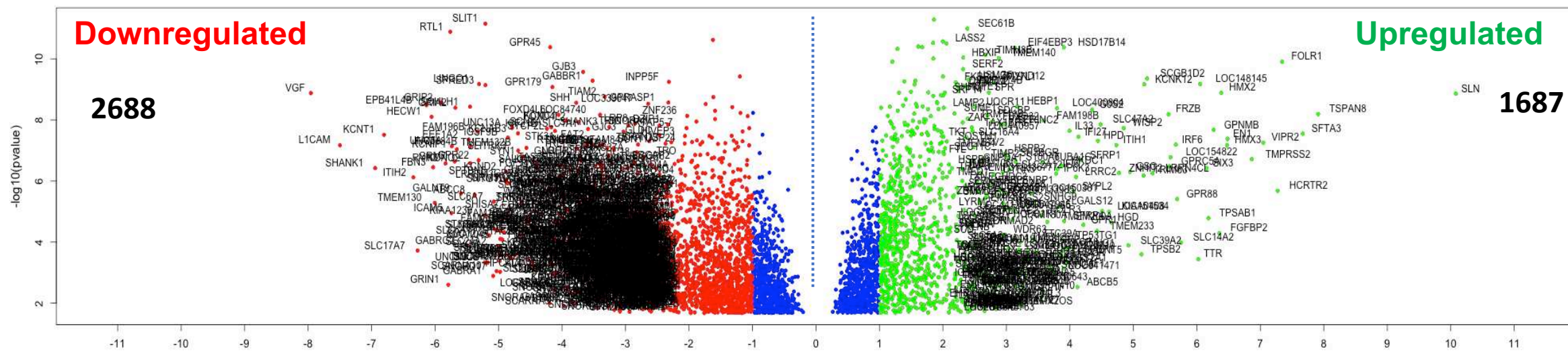


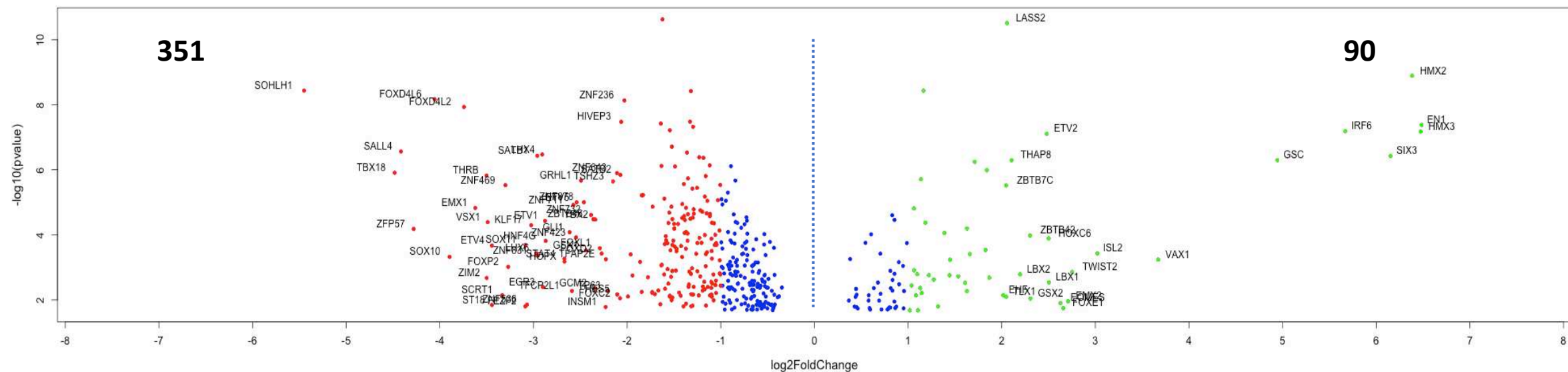
Fig. S3b

FDR<0.05

Distribution of all differentially expressed genes in SEGA and gangliogliomas



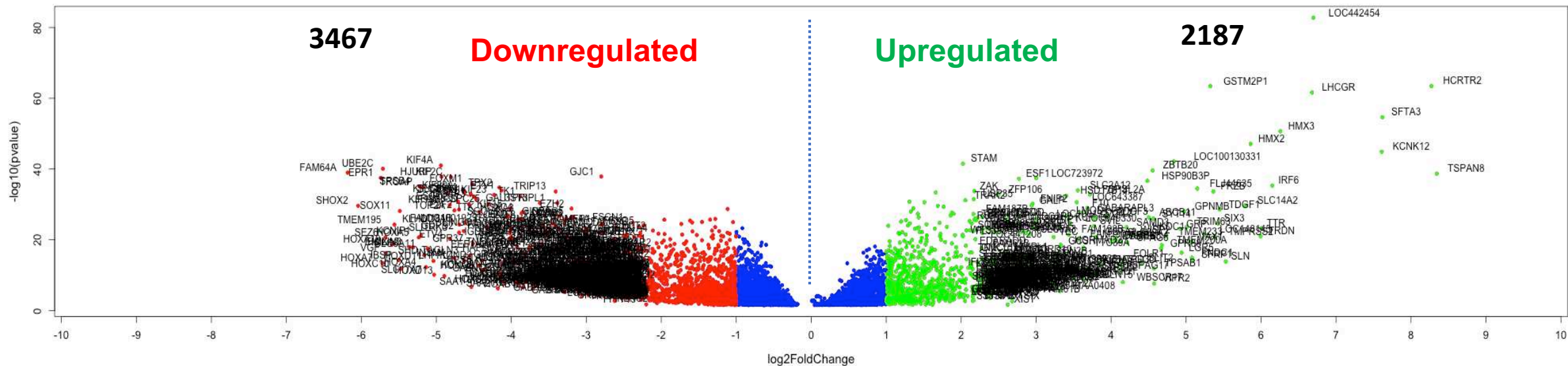
Distribution of all differentially expressed TFs in SEGA and gangliogliomas



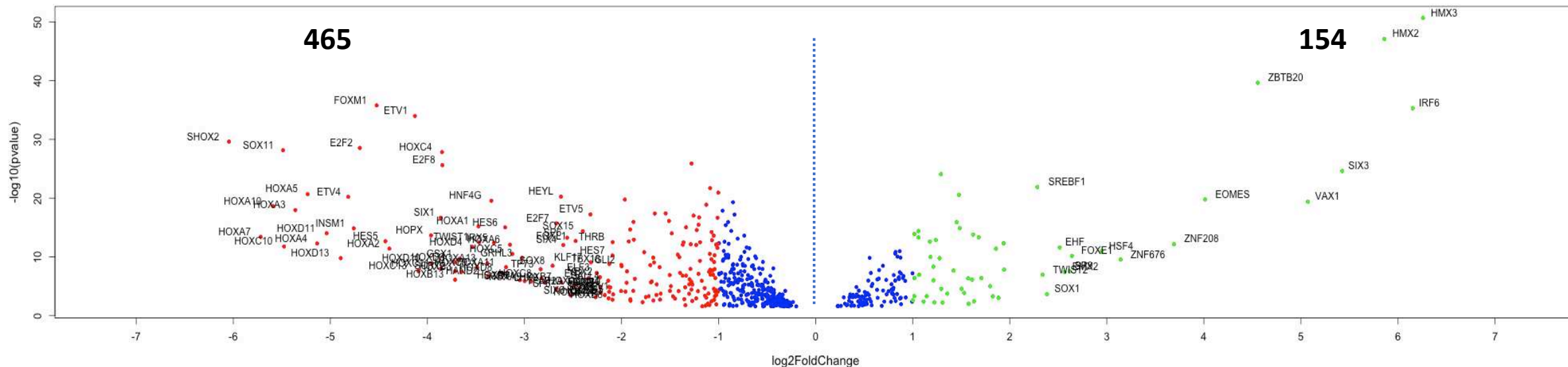
FDR<0.05

Fig. S3d

Distribution of all differentially expressed genes in SEGA and TCGA GBM tumors



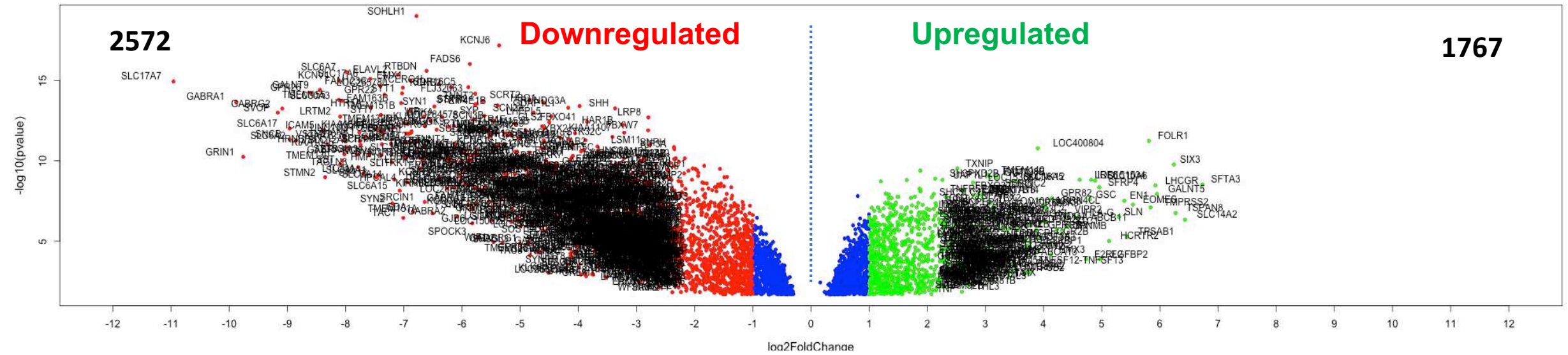
Distribution of all differentially expressed TFs in SEGA and TCGA GBM tumors



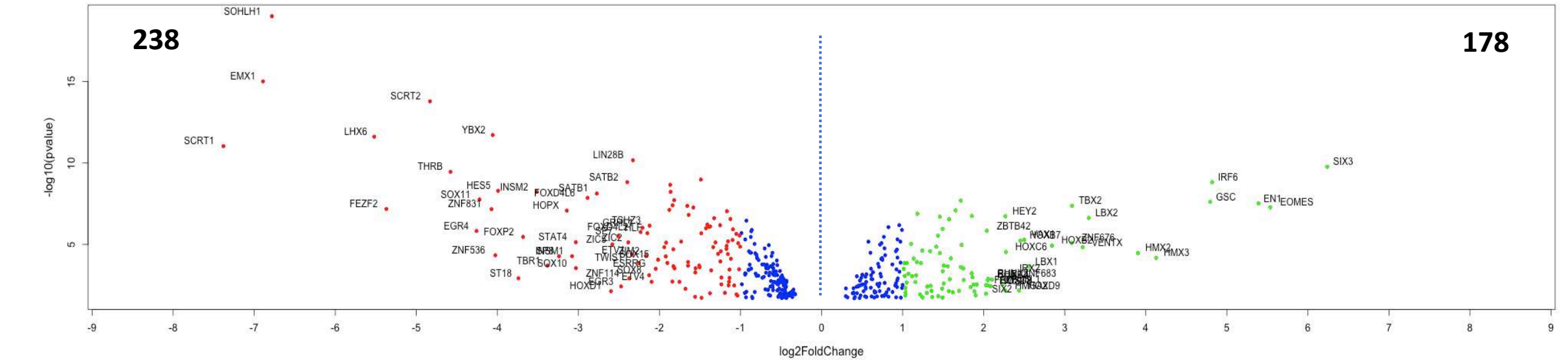
FDR<0.05

Fig. S3c

Distribution of all differentially expressed genes in SEGA and cortical tubers



Distribution of all differentially expressed TFs in SEGA and cortical tubers



FDR<0.05

Fig. S3e

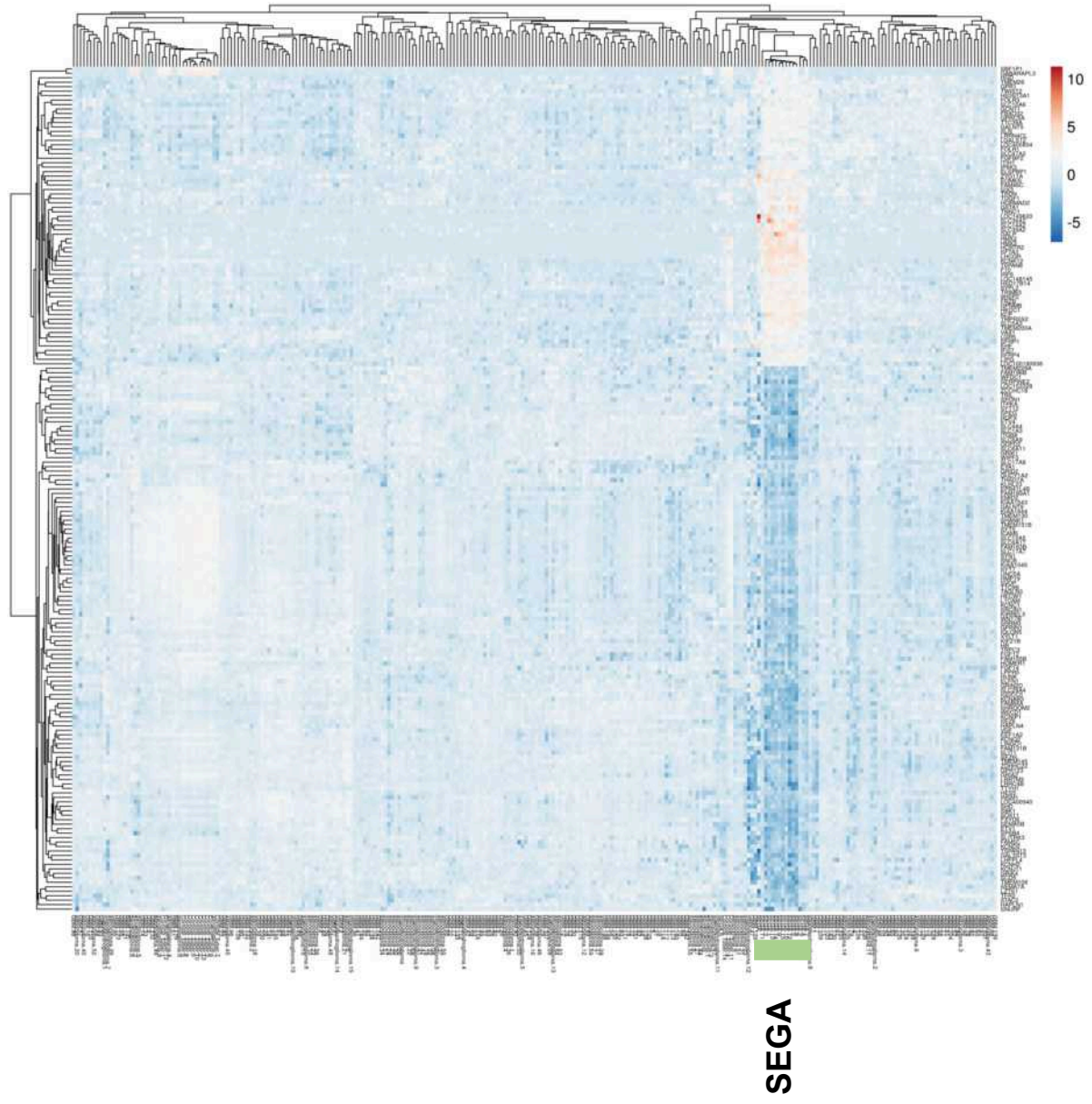
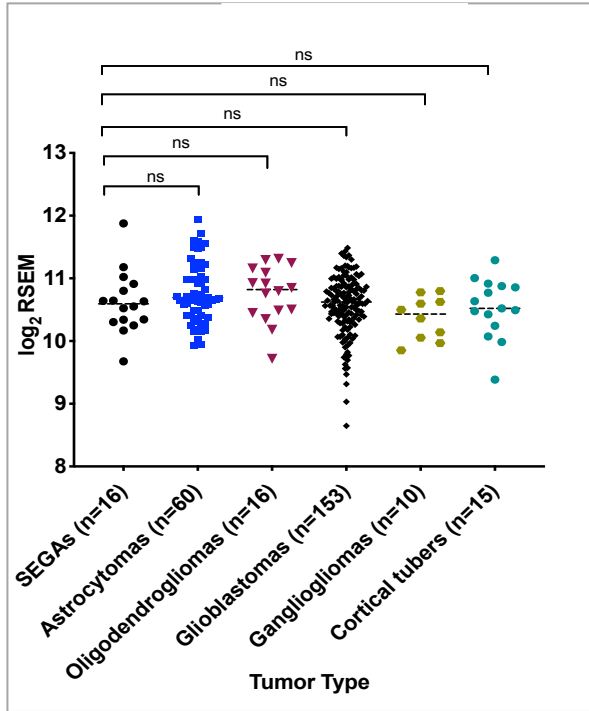
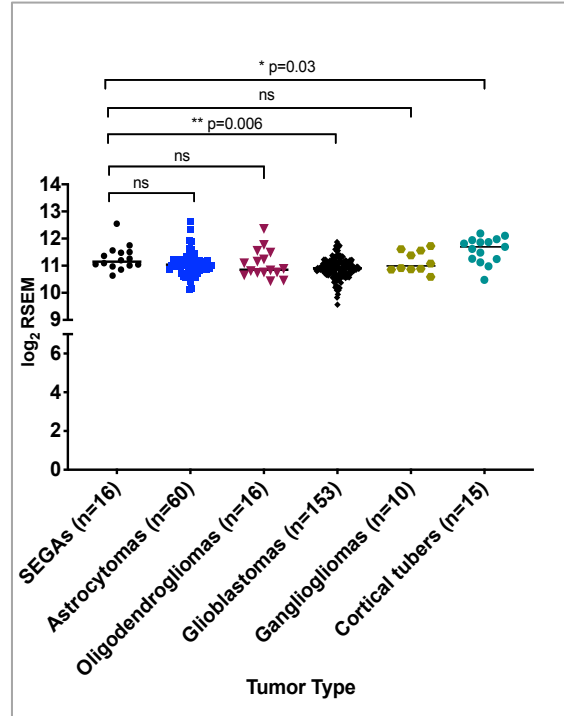


Fig. S3f

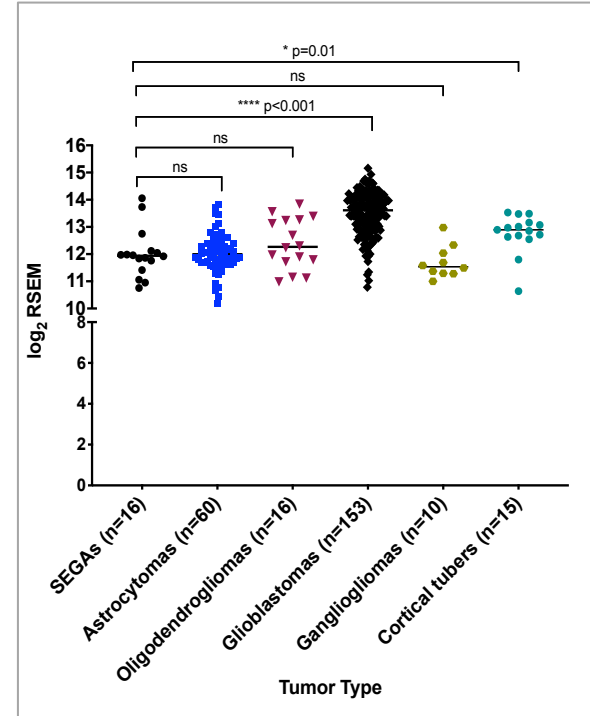
ABCF1



SDHA



LDHA



GAPDH

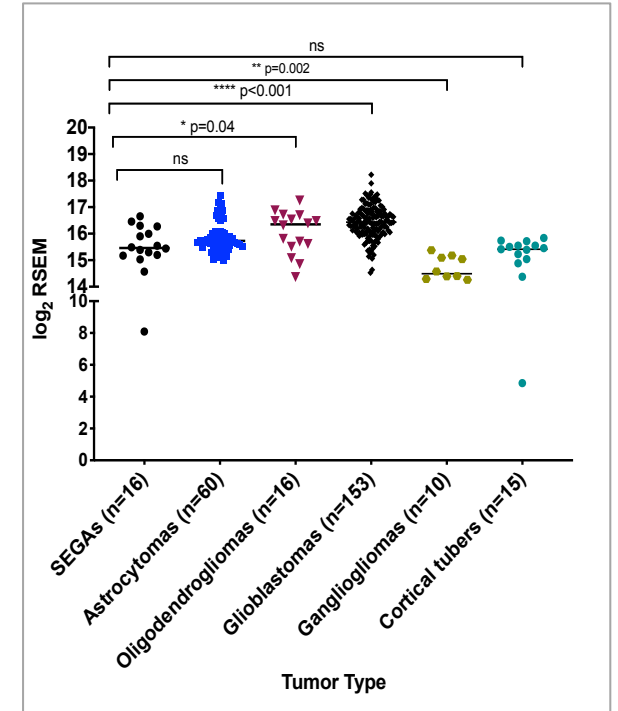
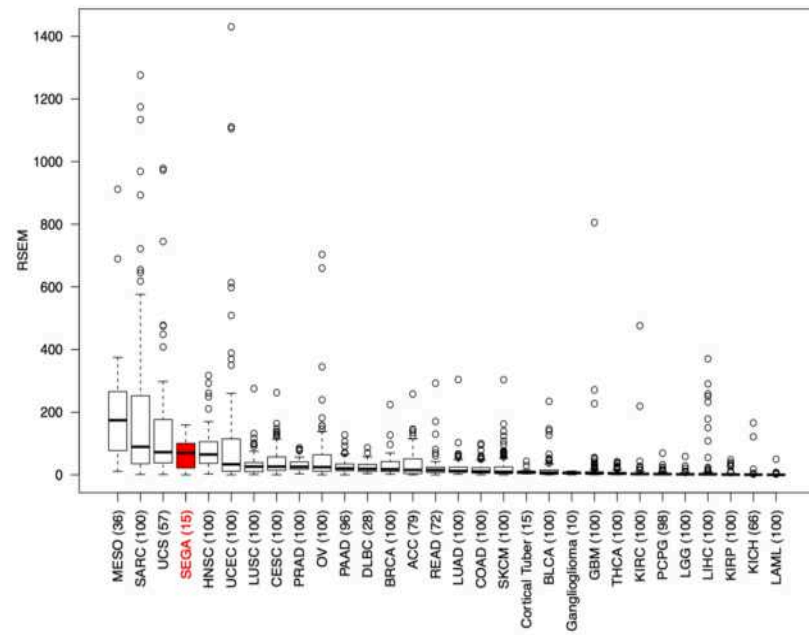


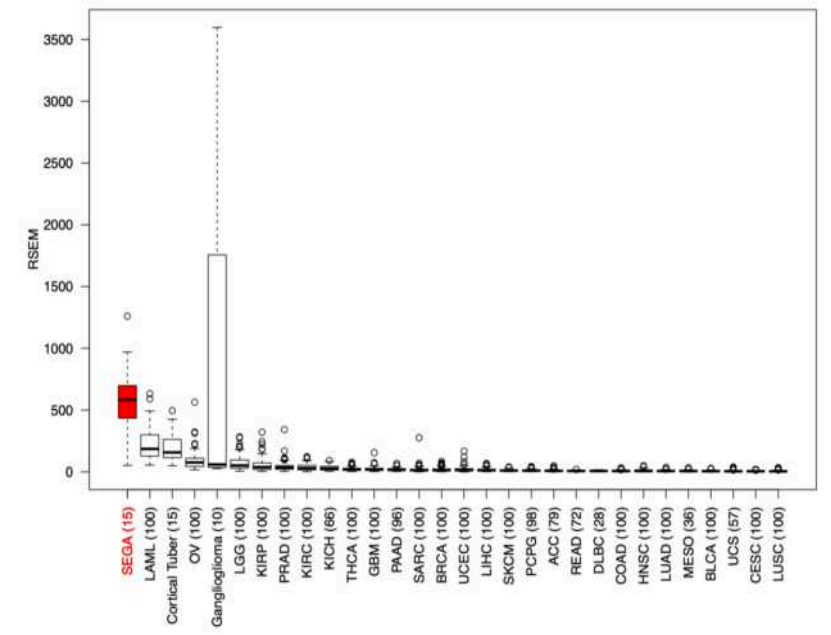
Fig. S4

TWIST2

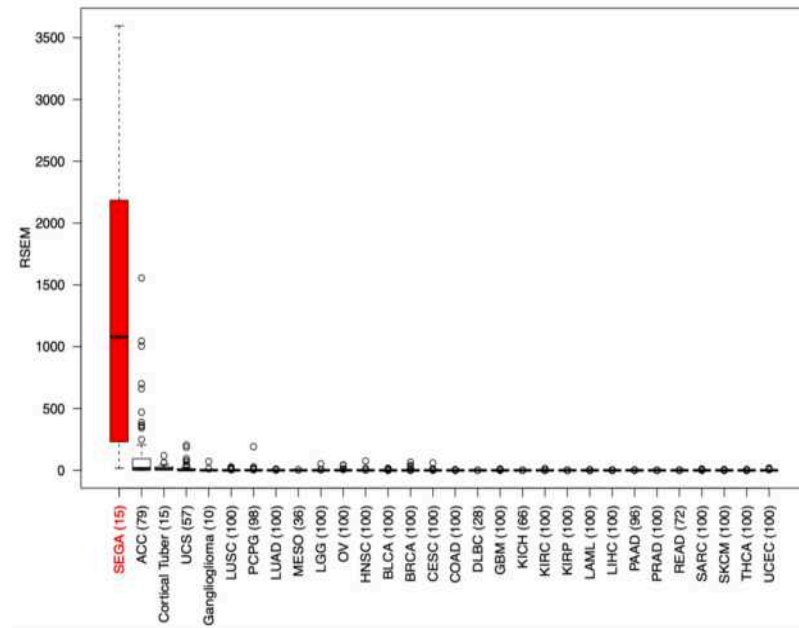


EHF

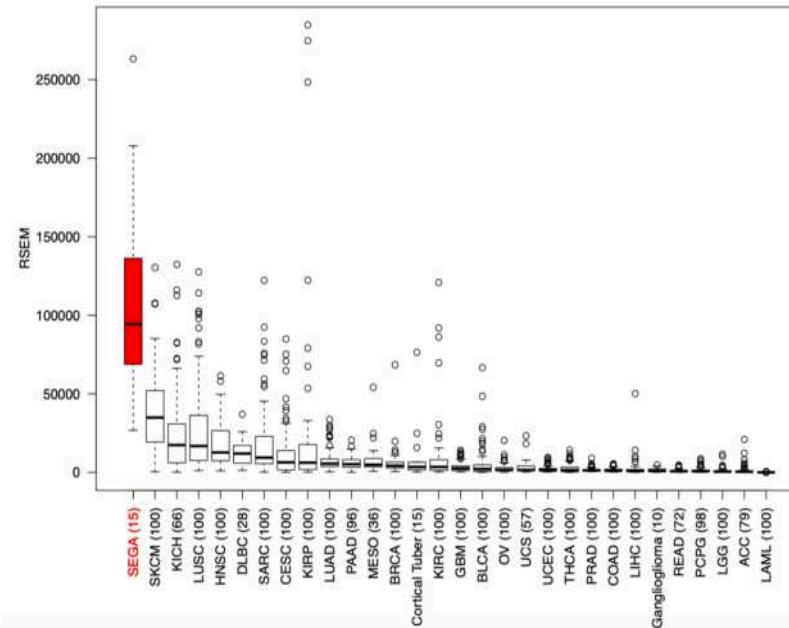
ZBTB20



HCRTR2



GPMNB



CTSK

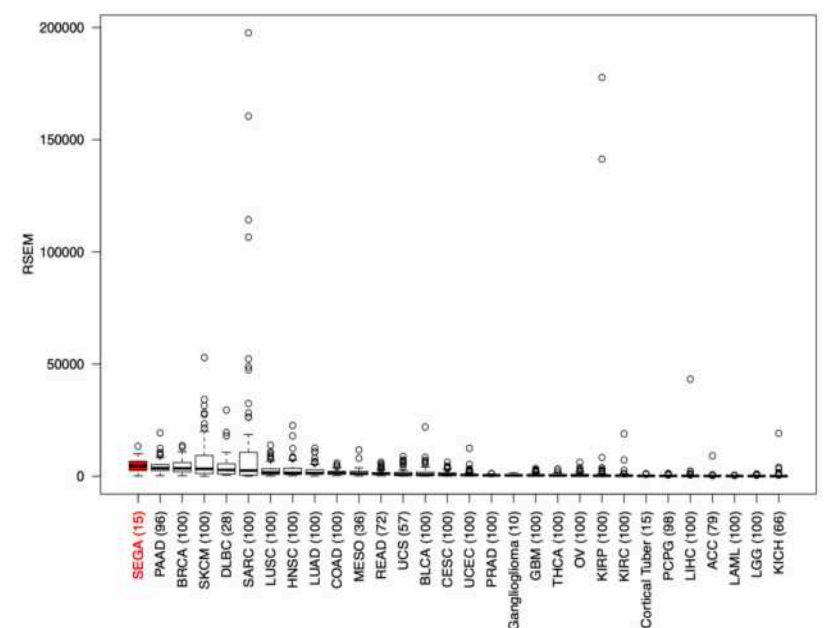
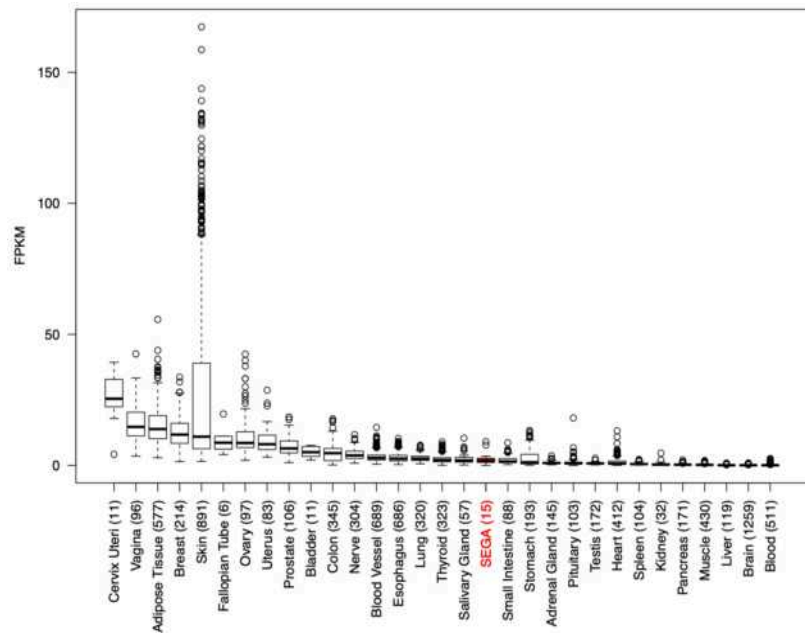
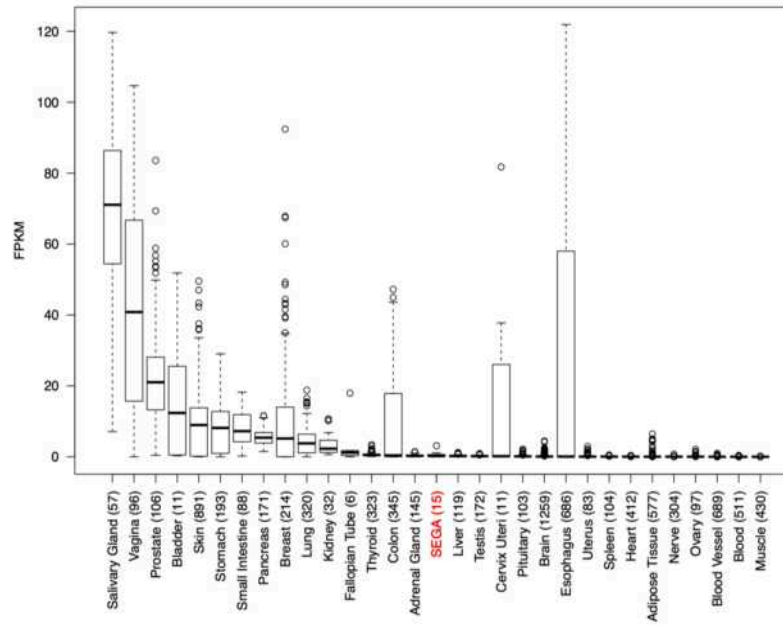


Fig. S5

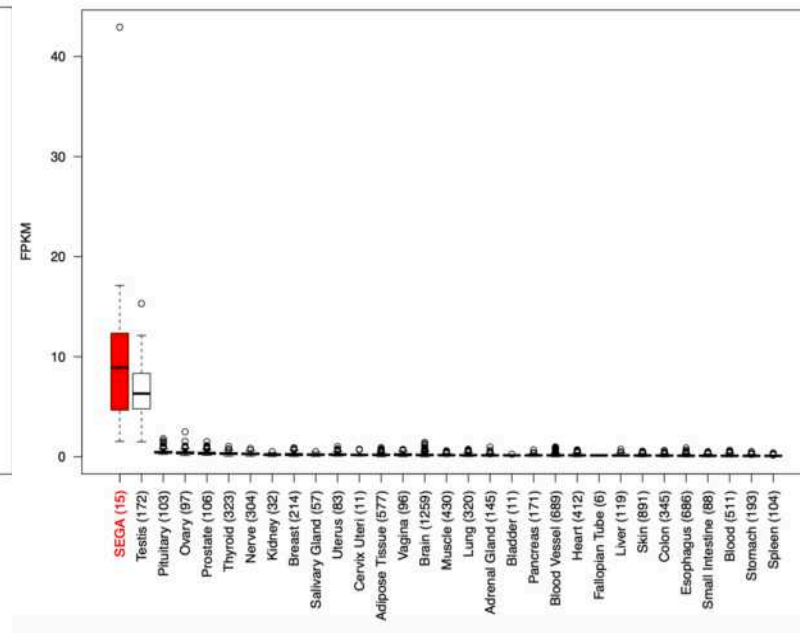
TWIST2



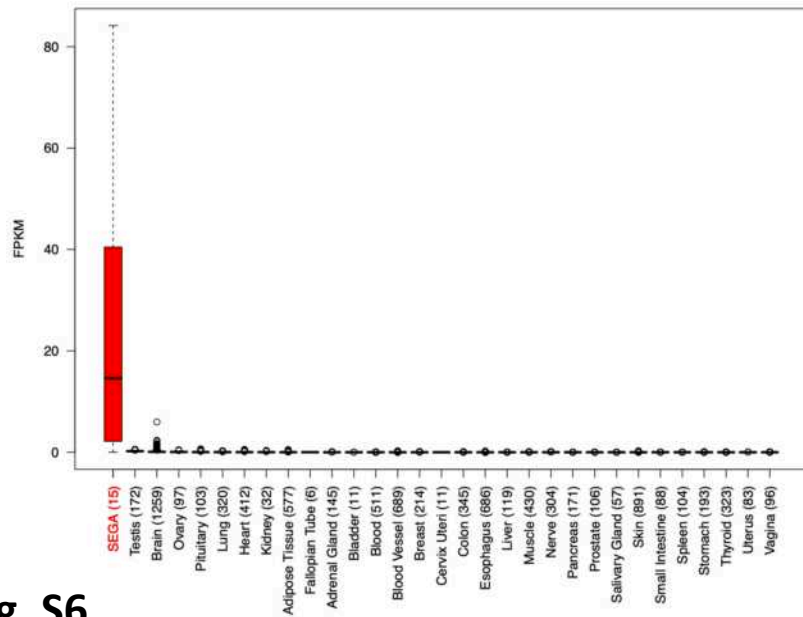
EHF



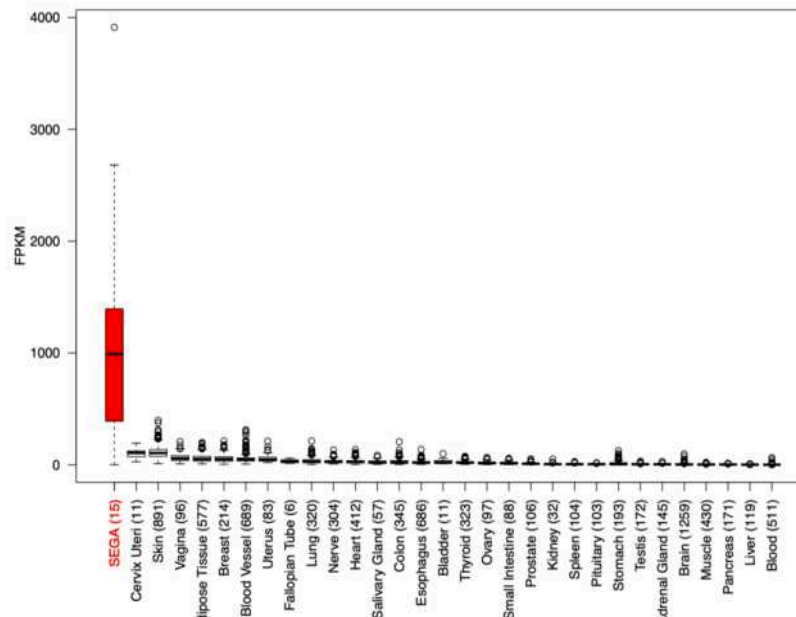
ZBTB20



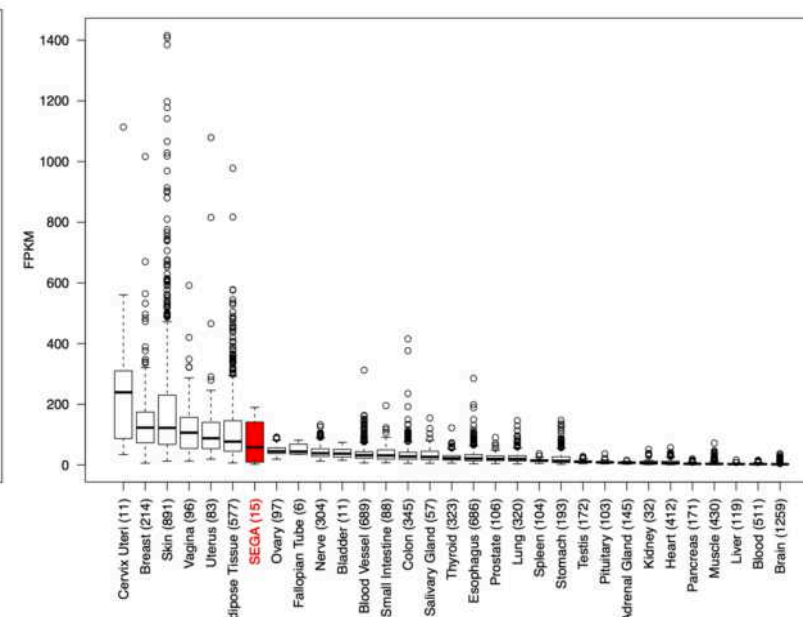
HCRTR2



GPMNB



CTSK



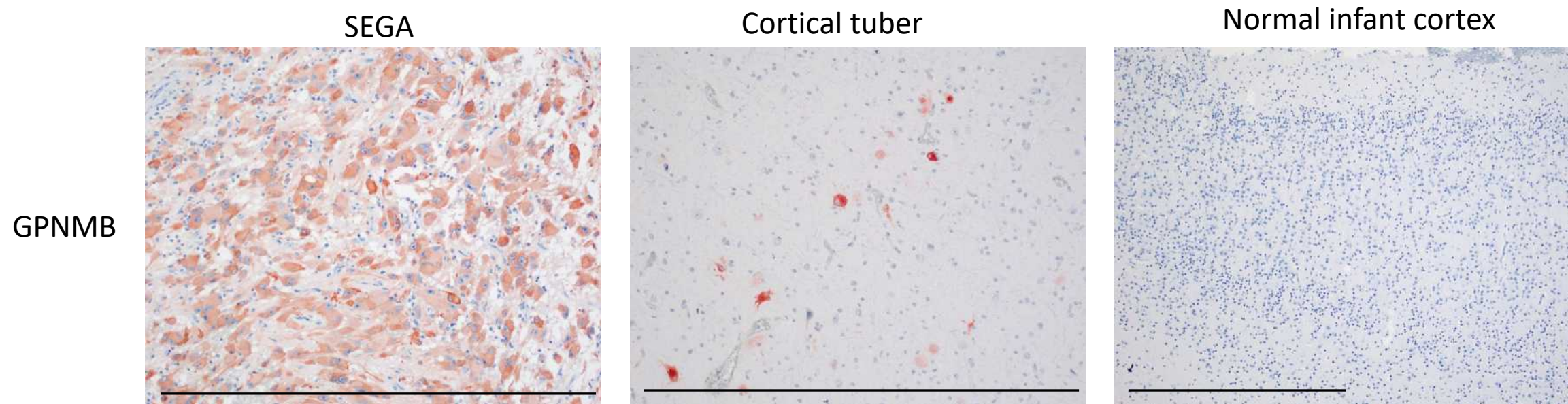
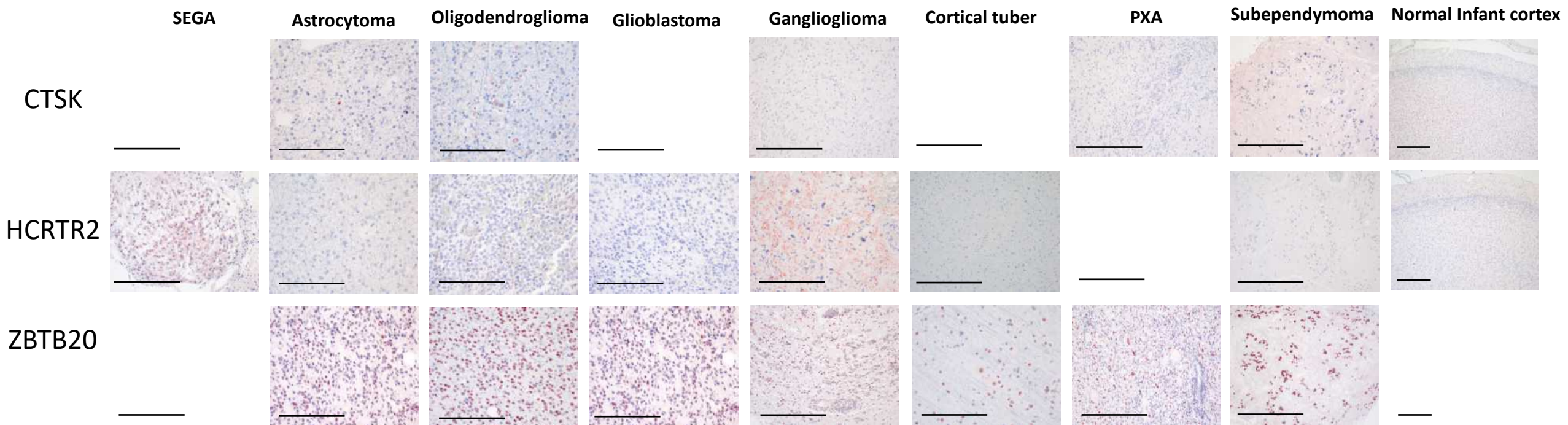


Fig. S7

