

# A (GCC) Repeat in the Untranslated Region of Human SBF1 Departs from Hardy-Weinberg Equilibrium in Human and Links to Late-onset Neurocognitive Disorder.

Mina Ohadi (✉ [ohadi.mina@yahoo.com](mailto:ohadi.mina@yahoo.com))

University of Social Welfare and Rehabilitation Sciences

**Safoura Khamse**

University of Social Welfare and Rehabilitation Sciences

**Samira Alizadeh**

University of Social Welfare and Rehabilitation Sciences

**Stephan H Bernhart**

IZBI, Universität Leipzig

**Hossein Afshar**

University of Social Welfare and Rehabilitation Sciences

**Ahmad Delbari**

University of Social Welfare and Rehabilitation Sciences

---

## Research Article

**Keywords:** SBF1, (GCC) repeat, phenomenon, short tandem repeat, purifying selection, late-onset NCD

**Posted Date:** January 13th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1240759/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The human *SBF1* (SET binding factor 1) gene, alternatively known as *MTMR5*, is predominantly expressed in the brain, and its epigenetic dysregulation is linked to late-onset neurocognitive disorders (NCDs), such as Alzheimer's disease. This gene contains a (GCC)-repeat at the interval between +1 and +60 of the transcription start site (SBF1-202 ENST00000380817.8). Sequencing of the *SBF1* (GCC)-repeat in a sample of 542 Iranian individuals, consisting of late-onset NCDs (N=260) and controls (N=282) revealed a predominantly bi-allelic locus for this STR, consisting of 8 and 9 repeats, with allele frequencies ranging from 0.39 to 0.55, and four other alleles with frequencies of <0.03 across the two groups. Overall heterozygosity for the observed alleles was significantly less than expected in the NCD and control groups, at 22.3% and 16.31%, respectively ( $p=0.000$ ). Specifically, the heterozygous 8/9 genotype was significantly less than expected in both case and control groups (Hardy-Weinberg disequilibrium,  $p=0.000$ ), and significantly enriched in the NCD group (Yates corrected  $p=0.001$ ). Skewed heterozygous genotypes were also detected for other allele combinations, such as 6/8 vs 6/9 across groups ( $p=0.000$ ). Bioinformatics studies revealed that the number of (GCC)-repeats may change the RNA secondary structure and interaction sites across human exon 1. This STR was specifically expanded beyond 2-repeats in primates. In conclusion, we report a novel biological phenomenon in which there is indication of purifying selection against heterozygous genotypes at a STR locus in human, and skewed genotype compartment in late-onset NCD vs. controls. In view of the location of this STR in the 5' UTR, RNA/RNA or RNA/DNA heterodimer formation of the involved genotypes and possible deleterious downstream events should be considered.

## Introduction

Despite their vast evolutionary and biological implications ([1–8], short tandem repeats (STRs) remain an underappreciated topic in comparison to single nucleotide substitutions [9, 10], partly because of their repetitive nature and hardship of accurate allele calling with the currently available methods.

Among various categories of STRs, CGG/GCC repeats are overrepresented in the exons of the human genome, and are mainly focused on because of their involvement in neurological disorders [11–14]. The human gene, *SBF1* (SET binding factor 1), also known as *MTMR5* (Myotubularin-related protein 5) contains an annotated (GCC)-repeat of 9-repeats in the interval between +1 to +60 of the transcription start site (TSS) (SBF1-202 ENST00000380817.8), which is in the top 1 percentile of (GCC)-repeats with respect to length [15]. *SBF1* is located at the extreme end of the long arm of chromosome 22 (22q13.33), and across all human tissues, reaches maximum expression in the human cortex (<https://www.proteinatlas.org/ENSG00000100241-SBF1/tissue>). In comparison with other primate species, *SBF1* reaches maximum expression quantiles in the human brain and skeletal muscle (<https://www.ncbi.nlm.nih.gov/IEB/Research/Acembly>) [16]. In line with the above, aberrant regulation of the gene networks in which *SBF1* plays a role has been reported in late-onset neurocognitive disorders (NCDs), such as Alzheimer's disease (AD) [17].

Here we sequenced the *SBF1* (GCC)-repeat in a sample of humans, consisting of late-onset NCDs and controls. We also studied the status of this STR across vertebrates.

## Materials And Methods

### Subjects

Five hundred forty-two unrelated Iranian subjects of  $\geq 60$  years of age, consisting of late-onset NCD patients (DSM-5) (N=260) and controls (N=282) were recruited from the provinces of Tehran, Qazvin, and Rasht. In each NCD case, the Persian version of the Abbreviated Mental Test Score (AMTS) [18, 19] was implemented (AMTS $<7$  was an inclusion criterion for NCD), medical records were reviewed in all participants, and CT-scans were taken where possible. Furthermore, in a number of subjects, the Mini-Mental State Exam (MMSE) Test [20] was implemented in addition to the AMTS. A score of  $<24$  was an inclusion criterion for NCD. The Persian version of the AMTS is a valid cognitive assessment tool for older Iranian adults, and can be used for NCD screening in Iran [18]. The control group was selected based on cognitive AMTS of  $>7$  and MMSE $>24$ , lack of major medical history, and normal CT-scan where possible. The cases and controls were matched based on age, gender, and residential district. The subjects' informed consent was obtained (from their guardians where necessary) and their identities remained confidential throughout the study. The research was approved by the Ethics Committee of the Social Welfare and Rehabilitation Sciences, Tehran, Iran, and was consistent with the principles outlined in an internationally recognized standard for the ethical conduct of human research. All methods were performed in accordance with the relevant guidelines and regulations.

### Allele and genotype analysis of the *SBF1* (GCC)-repeat.

Genomic DNA was obtained from peripheral blood using a standard salting out method. PCR reactions for the amplification of the *SBF1* (GCC)-repeat were set up with the following primers:

**Forward:** TCTGGACCAATGGAGATGCG

**Reverse:** GAAGTAGTCCGCGAGCCG

PCR reactions were carried out in a final volume of 20  $\mu$ l, at a final concentration of 30% high-GC buffer, in a thermocycler (PeqLab-PEQStar) under the following conditions: initial denaturation at 95 °C for 5 min, 40 cycles of denaturation at 95 °C for 45 s, annealing at 55 °C for 45 s, and extension at 72 °C for 1 min, and a final extension at 72 °C for 10 min. All samples included in this study were sequenced by the forward primer, using an ABI 3130 DNA sequencer.

### Statistical analysis

The OpenEpi software (<https://www.openepi.com/TwoByTwo/TwoByTwo.htm>) was implemented to analyze the allele and genotype data in the human samples studied.

# Structural analysis of the human SBF1 with different numbers of (GCC)-repeats

We investigated accessibility (probability to be unpaired) differences of exon 1 of the human *SBF1* gene with five to nine (GCC)-repeats, using the accessibility computation of the ViennaRNA package (RNAplfold with `-W 300 -L 300 -u 10`) [22, 23]. We compared the accessibilities of all regions of 10 nt length, and found distinct differences at 3 regions of exon 1. Furthermore, we used RNAup -b [24] to compare possible interactions in homodimeric and heterodimeric *SBF1* first exon with different numbers of (GCC)-repeats.

## Analysis of the SBF1 (GCC)-repeat across vertebrates.

The interval between +1 and +100 of the TSS of the *SBF1* was searched across all species in which *SBF1* was annotated, based on Ensembl 104. The Ensembl alignment program was implemented for the sequence alignments across the selected species.

## Results

### The SBF1 (GCC)-repeat allele frequency compartment was significantly skewed in the NCD group vs. controls.

We detected two predominantly abundant alleles of 8 and 9-repeats in the human subjects studied in both groups (Table 1, Fig. 1). At significantly lower frequencies, we detected repeats of 5, 6, 7, and 10, with frequencies of  $<0.03$ . The allele frequency compartment was significantly skewed in the NCD group vs. controls, mainly due to the 8 and 9 repeat alleles (Yates corrected (Chi)<sup>2</sup> for 8-repeat=9.54,  $p=0.001$ ; Yates corrected (Chi)<sup>2</sup> for 9-repeat=6.29,  $p=0.006$ ).

Table 1  
Alleles and genotypes of the human *SBF1* (GCC) repeat in the NCD and control groups.

Allele count		
Allele (repeat)	Controls	NCDs
5	0	1
6	16	12
7	1	0
8	224	256
9	313	248
10	10	3
Genotype count		
Genotype	Controls	NCDs
5/6	0	1
6/8	12	11
6/9	4	0
7/8	1	0
8/8	93	100
8/9	23	45
8/10	2	0
9/9	141	101
9/10	4	1
10/10	2	1

## The *SBF1* (GCC)-repeat genotype compartment was significantly anomalous across the NCD and control groups.

Overall heterozygosity for the observed alleles was significantly less than expected in the NCD and control groups, at 22.3% and 16.31%, respectively ( $p=0.000$ ). Specifically, rather than an expected >45% 8/9 genotype based on the predominantly bi-allelic 8 and 9-repeat allele frequencies, we detected <18% of that genotype across the two groups ( $p=0.000$ ) (Table 1, Fig. 2). There were other discrepancies in the genotype distribution of alleles across the two groups. The 6/8 genotype was significantly more detected

than the 6/9 genotype in both case and control groups ( $p < 0.05$ ), and the skew was even more significant when the two groups were pooled ( $p < 0.0002$ ). Those discrepancies were detected for other heterozygous genotypes, such as the 9/10 and 8/10 genotypes, in which there was skew of the homo/hetero ratio in the observed vs. expected compartments in both groups in the context of the 10/10 homozygotes ( $p < 0.000$ ).

Between the two groups, we detected significant skew, primarily due to the significant enrichment of the 8/9 genotype in the NCD group vs. controls, and reverse ratio of 8/8 and 9/9 genotypes ( $p = 0.001$ ), which was primarily a consequence of excess 8/9 genotypes in the former.

## Identification of an extreme genotype in the NCD group only.

We detected a genotype at the extreme short end of the allele range in one instance of late-onset NCD. This genotype was 5/6 (Fig. 3), and was detected in an 85-year-old female case of NCD with AMTS=3, and suspected of having late-onset AD. Whereas the 5-repeat was the shortest allele detected in the NCD group, the shortest allele detected in the control group was 6-repeats.

## The number of (GCC)-repeats may change the RNA secondary structure and interaction sites.

The accessibility of exon 1 of *SBF1* varied with the number of (GCC)-repeats in three regions, around nt 50 (at the (GCC)-repeat itself), at about nt 200 (at the translation start site) and at nt 220 (all nt relative to the (GCC)<sub>9</sub> variant based on Ensembl transcript ID: ENST00000380817.8 SBF1-202) (Fig. 4).

Furthermore, we analyzed where the preferred interaction sites between two *SBF1* exons 1 and 2 would be, and found that there are two different groups of interaction sites (Table 2): in one group, the best molecular interaction occurs between nt 119-130 and nt 219-230, while the other group has interactions between nt 182-200 and nt 193-211 (all nt relative to the (GCC)<sub>9</sub>).

Table 2  
Interaction groups across various human *SBF1* (GCC)-repeats\*

<p>(((((((((&amp;)))))).))))))</p> <p><b>CGUGCUGGUGGC&amp;GCCAUGAGCGCG</b></p> <p><b>5 vs 5, 5 vs 6, 5 vs 7, 5 vs 8, 5 vs 9, 6 vs 6, 6 vs 7, 6 vs 8, 8 vs 8</b></p>
<p>(((((((((((((&amp;)))))).)).))))))</p> <p><b>GCCAUGGCGCGGCUCGCGG&amp;CCGCGUCCCUCGCCAUGGC</b></p> <p><b>6 vs 9, 7 vs 7, 7 vs 8, 7 vs 9, 8 vs 9, 9 vs 9</b></p>
<p>*Lengths with interaction structure and sequences in bracket-dot notation</p>
<p>(matching parenthesis are opening and closing bases of a base pair, dots</p>
<p>are unpaired bases, and separate the two interacting sequences).</p>

<p>(((((((((&amp;)))))).))))))</p> <p><b>CGUGCUGGUGGC&amp;GCCAUGAGCGCG</b></p> <p><b>5 vs 5, 5 vs 6, 5 vs 7, 5 vs 8, 5 vs 9, 6 vs 6, 6 vs 7, 6 vs 8, 8 vs 8</b></p>
<p>(((((((((((((&amp;)))))).)).))))))</p> <p><b>GCCAUGGCGCGGCUCGCGG&amp;CCGCGUCCCUCGCCAUGGC</b></p> <p><b>6 vs 9, 7 vs 7, 7 vs 8, 7 vs 9, 8 vs 9, 9 vs 9</b></p>

## **SBF1 (GCC)-repeat expanded specifically in primates.**

Across all the vertebrate species studied, the *SBF1* (GCC)-repeat specifically expanded beyond 2-repeats in primates (Fig. 5).

## **Discussion**

Here we report the first indication of purifying selection at a STR locus in human. The primary importance of (GCC)-repeats stems from a possible link between that type of STR and natural selection, mainly for

two reasons: Firstly, (GCC)-repeats are specifically enriched in the exons. Secondly, GC-rich sequences are mutation hotspots [25], and frequently interrupted by single nucleotide substitutions. Specific expansion of the *SBF1* (GCC)-repeat in primates, and not in any other order, supports selective advantage of the STR in this order.

In both NCD and control groups, the expected heterozygosity for the observed allele frequencies was dramatically compromised, most likely due to selection against heterozygous genotypes. As a consequence, the homozygote compartment expanded significantly beyond expectation and over 77% across the two groups. This anomaly could not be attributed to the excess of consanguineous marriages in Iran, as excess of homozygosity in consanguineous societies can contribute to between 2 and 11% homozygosity at a given locus [26, 27]. The homozygous genotypes could not be attributed to allele dropout either, as the frequency of such event is less than 0.004 in amplification-based approaches [28]. Sampling error is another explanation for the observed genotypes. All samples were collected from the same districts in Iran, and the results were replicated in both groups, such as the shrunk 8/9 genotype compartment, and the excess of the 6/8 vs. 6/9 genotypes. However, it should be noted that this is a pilot study, and warrants replication by independent studies.

Searching the Genome Aggregation database (gnomAD) for the human *SBF1* (GCC)-repeat revealed inconclusive data for the annotated alleles and genotypes, which spanned across all the populations studied (<https://gnomad.broadinstitute.org>). The above findings are most likely due to the frequent failure of the general whole-exome sequencing methods to capture GC-rich sequences. Successful PCR amplification of the human *SBF1* gene necessitates stringent conditions and special GC-rich buffer preparations as described in the Methods.

A likely hypothesis that may be put forward is that the heterozygous genotypes might have been selected against in human in the process of evolution. The studied (GCC)-repeat is located in the 5' UTR, and it may be speculated that the heterodimer RNAs of, for example, 8 and 9 repeats, and 6 and 9, have a detrimental effect on the downstream events, such as transcript processing and translation. A possible mechanism might be connected to RNA structure and accessibility, which we could show does change with the number of (GCC)-repeats, and can affect at least exon 1.

Example of RNA heterodimer formation exists in the 5' regulatory regions of human HIV-1/HIV-2 RNAs [29, 30], which requires GC-rich palindromic sequences among a number of other motifs [31]. It may be speculated that similar sequences in the GC-rich human *SBF1* RNAs fulfill the conditions for potential RNA dimerization. Experimental synthetic stem-loop RNAs have been reported to alter the expression of a number of genes in bacteria [32].

*SBF1* is predominantly expressed in the brain and skeletal muscle, and the protein encoded by this gene is a member of the myotubularin family. Myotubularin-related proteins, namely MTMR2, MTMR13/SBF2 and MTMR5/SBF1 are mainly involved in regulating endolysosomal trafficking [33] and mitochondrial functioning [34]. Dysregulation of *SBF1* is linked to late-onset NCDs such as AD [17], which is also indicated by the observed genotype anomalies in the NCD group vs. controls in our study. An isolate

instance of an NCD patient harboring a genotype that consisted of extreme short alleles, may be of significance, while random co-occurrence should also be considered as a possibility. The secondary structure and accessibility effect of the 5/6 genotype were dramatically divergent, and the 5-repeat allele length was not detected in the control group. It is possible that low frequency alleles at the extreme ends of the allele distribution curve are subject to negative natural selection [8, 12, 35].

It remains to be clarified how certain heterozygous genotypes might have been selected against in human, and may increase the risk of late-onset AD. It is also warranted that this STR is sequenced in larger samples and in a spectrum of neurological disorders.

## Conclusion

We report indication of a novel biological phenomenon, in which there is significant selection against heterozygous genotypes at a STR locus in the human population. In view of the location of the (GCC)-repeat in the 5' UTR of the *SBF1* gene, it is speculated that specific RNA/RNA or DNA/RNA heterodimers may exert effects that are favored against in the course of evolution. We also report skewed genotypes in late-onset NCD vs. controls at this locus. It should be noted that this is a pilot study, warranting replication by independent studies.

## Abbreviations

AMTS: Abbreviated Mental Test Score

MTMR5: Myotubularin-Related Protein 5

nt: Nucleotide

SBF1: SET binding factor 1

STR: Short tandem repeat

TSS: Transcription start site

UTR: Untranslated region

## Declarations

## Author Contributions Statement

Safoura Khamse and Samira Alizadeh performed the molecular experiments and allele and genotype analyses. Stephan H Bernhart performed the RNA structural and accessibility analyses. Hossein Afshar collected the human samples and their clinical information. Ahmad Delbari contributed to data collection and coordination. Mina Ohadi conceived and supervised the project, and wrote the manuscript.

## References

1. Nikkhah, M., et al., *An exceptionally long CA-repeat in the core promoter of SCGB2B2 links with the evolution of apes and Old World monkeys*. *Gene*, 2016. **576**(1 Pt 1): p. 109–14.
2. Mohammadparast, S., et al., *Exceptional expansion and conservation of a CT-repeat complex in the core promoter of PAXBP1 in primates*. *Am J Primatol*, 2014. **76**(8): p. 747–56.
3. Afshar, H., et al., *Evolving evidence on a link between the ZMYM3 exceptionally long GA-STR and human cognition*. *Scientific Reports*, 2020. **10**.
4. Watts, P., et al., *Stabilizing selection on microsatellite allele length at arginine vasopressin 1a receptor and oxytocin receptor loci*. *Proceedings of the Royal Society B: Biological Sciences*, 2017. **284**: p. 20171896.
5. Hannan, A.J., *Tandem repeats mediating genetic plasticity in health and disease*. *Nature Reviews Genetics*, 2018. **19**(5): p. 286–298.
6. Ranathunge, C., et al., *Microsatellites as Agents of Adaptive Change: An RNA-Seq-Based Comparative Study of Transcriptomes from Five Helianthus Species*. *Symmetry*, 2021. **13**(6): p. 933.
7. Fotsing, S.F., et al., *The impact of short tandem repeat variation on gene expression*. *Nat Genet*, 2019. **51**(11): p. 1652–1659.
8. Afshar, H., et al., *Natural Selection at the NHLH2 Core Promoter Exceptionally Long CA-Repeat in Human and Disease-Only Genotypes in Late-Onset Neurocognitive Disorder*. *Gerontology*, 2020. **66**(5): p. 514–522.
9. Press, M.O., et al., *Substitutions Are Boring: Some Arguments about Parallel Mutations and High Mutation Rates*. *Trends in genetics: TIG*, 2019. **35**(4): p. 253–264.
10. Press, M.O., K.D. Carlson, and C. Queitsch, *The overdue promise of short tandem repeat variation for heritability*. *Trends Genet*, 2014. **30**(11): p. 504–12.
11. Annear, D.J., et al., *Abundancy of polymorphic CGG repeats in the human genome suggest a broad involvement in neurological disease*. *Scientific Reports*, 2021. **11**(1): p. 2515.
12. Jafarian, Z., et al., *Natural selection at the RASGEF1C (GGC) repeat in human and divergent genotypes in late-onset neurocognitive disorder*. *Sci Rep*, 2021. **11**(1): p. 19235.
13. Khamse, S., et al., *Novel implications of a strictly monomorphic (GCC) repeat in the human PRKACB gene*. *Scientific reports*, 2021. **11**(1): p. 20629–20629.
14. Tang, H., et al., *Profiling of Short-Tandem-Repeat Disease Alleles in 12,632 Human Whole Genomes*. *Am J Hum Genet*, 2017. **101**(5): p. 700–715.
15. Namdar, P., et al., *Exceptionally long 5' UTR short tandem repeats specifically linked to primates*. *Gene*, 2015. **569**.
16. Thierry-Mieg, D. and J. Thierry-Mieg, *AceView: a comprehensive cDNA-supported gene and transcripts annotation*. *Genome Biol*, 2006. **7 Suppl 1**(Suppl 1): p. S12.1–14.

17. Li, P., et al., *Epigenetic dysregulation of enhancers in neurons is associated with Alzheimer's disease pathology and cognitive symptoms*. Nat Commun, 2019. **10**(1): p. 2246.
18. Foroughan, M., et al., *Validity and reliability of a abbreviated Mental Test Score (AMTS) among older Iranian*. Psychogeriatrics, 2017. **17**(6): p. 460–465.
19. Hodkinson, H.M., *Evaluation of a mental test score for assessment of mental impairment in the elderly*. Age Ageing, 1972. **1**(4): p. 233–8.
20. Carpenter, C.R., et al., *Accuracy of Dementia Screening Instruments in Emergency Medicine: A Diagnostic Meta-analysis*. Acad Emerg Med, 2019. **26**(2): p. 226–245.
21. Bernhart, S.H., U. Mückstein, and I.L. Hofacker, *RNA Accessibility in cubic time*. Algorithms for Molecular Biology, 2011. **6**(1): p. 3.
22. Lorenz, R., et al., *ViennaRNA Package 2.0*. Algorithms for Molecular Biology, 2011. **6**(1): p. 26.
23. Mückstein, U., et al., *Thermodynamics of RNA–RNA binding*. Bioinformatics, 2006. **22**(10): p. 1177–1182.
24. Nesta, A.V., D. Tafur, and C.R. Beck, *Hotspots of Human Mutation*. Trends Genet, 2021. **37**(8): p. 717–729.
25. Woods, C.G., et al., *Quantification of homozygosity in consanguineous individuals with autosomal recessive disease*. Am J Hum Genet, 2006. **78**(5): p. 889–896.
26. Kirin, M., et al., *Genomic runs of homozygosity record population history and consanguinity*. PLoS One, 2010. **5**(11): p. e13996.
27. Blais, J., et al., *Risk of Misdiagnosis Due to Allele Dropout and False-Positive PCR Artifacts in Molecular Diagnostics: Analysis of 30,769 Genotypes*. J Mol Diagn, 2015. **17**(5): p. 505–14.
28. Damgaard, C.K., et al., *RNA interactions in the 5' region of the HIV-1 genome*. J Mol Biol, 2004. **336**(2): p. 369–79.
29. Dirac, A.M.G., et al., *Requirements for RNA heterodimerization of the human immunodeficiency virus type 1 (HIV-1) and HIV-2 genomes*. J Gen Virol, 2002. **83**(Pt 10): p. 2533–2542.
30. Bou-Nader, C. and J. Zhang, *Structural Insights into RNA Dimerization: Motifs, Interfaces and Functions*. Molecules, 2020. **25**(12).
31. Paulus, M., M. Haslbeck, and M. Watzele, *RNA stem-loop enhanced expression of previously non-expressible genes*. Nucleic Acids Res, 2004. **32**(9): p. e78.
32. Mammel, A.E., et al., *Distinct roles for the Charcot-Marie-tooth disease-causing endosomal regulators Mtmr5 and Mtmr13 in axon radial sorting and Schwann cell myelination*. Human molecular genetics, 2021.
33. Berti, B., et al., *Bi-allelic variants in MTMR5/SBF1 cause Charcot-Marie-Tooth type 4B3 featuring mitochondrial dysfunction*. BMC medical genomics, 2021. **14**(1): p. 157–157.
34. Khamse et al. *Predominant monomorphism of the RIT2 and GPM6B exceptionally long GA blocks in human and enriched divergent alleles in the disease compartment*. Genetica, 2022. doi: 10.1007/s10709-021-00143-5. Online ahead of print. PMID: 34984576

# Figures

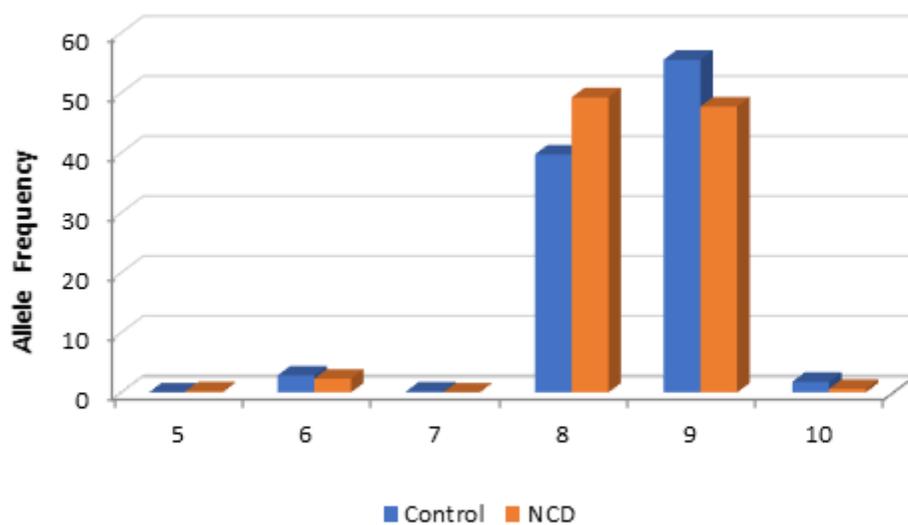
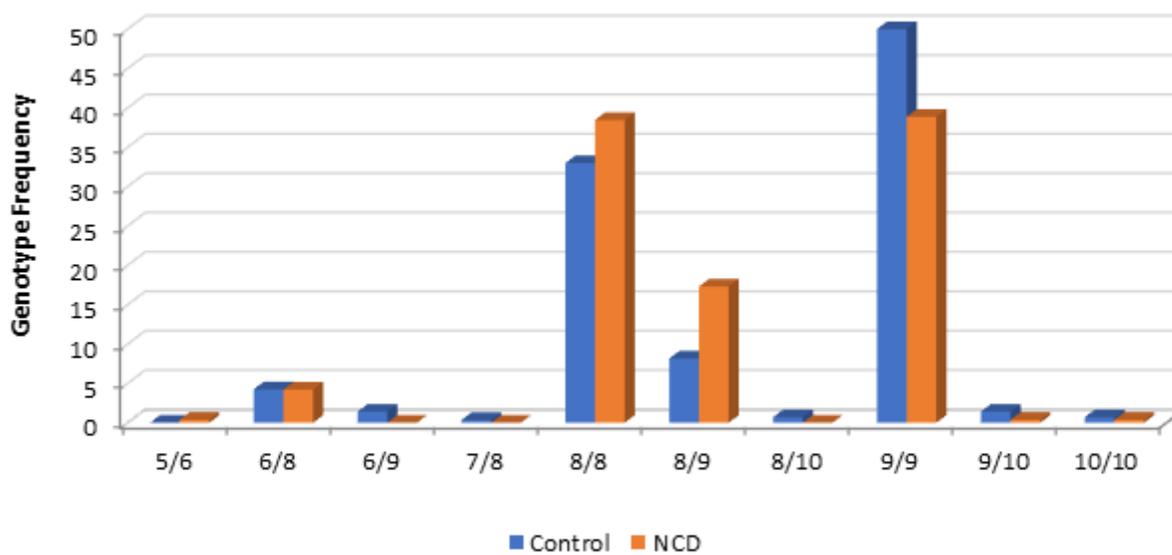
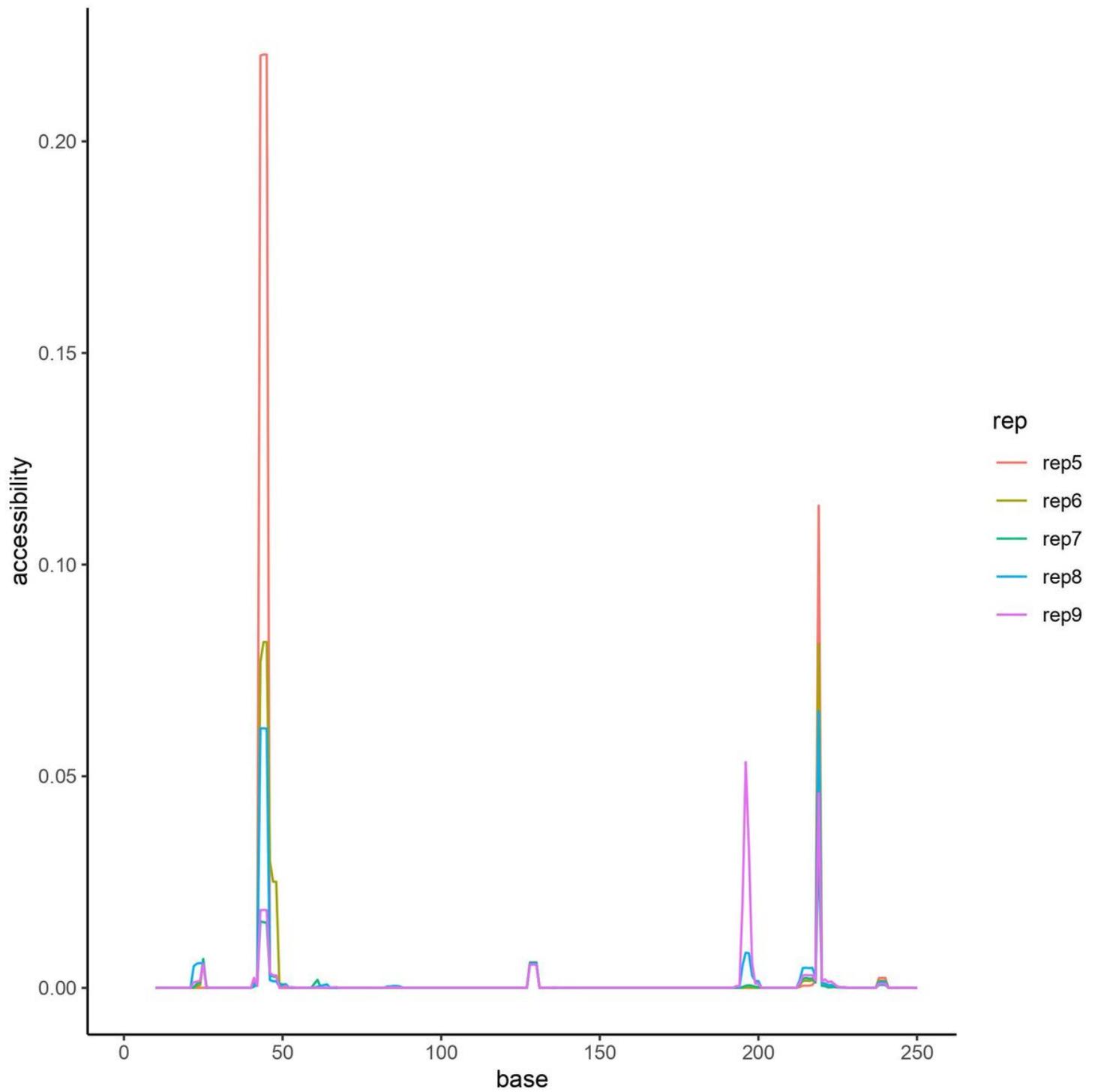


Figure 1

Allele frequency of the *SBF1* (GCC)-repeat in the human samples studied. The locus was predominantly biallelic, consisting of 8 and 9-repeat alleles.







**Figure 4**

Accessibility (probability to be unpaired) of all regions of 10 nt length, ending at base  $x$  for the first exon of human *SBF1* with 5 (red), 6 (green), 7 (black), 8 (blue), and 9 (black) (GCC) repeats. Differences in 3 regions were detected, at about nt 50, about nt 200, and about nt 220.

