Name: David Vandenbergh Email: djv4@psu.edu

Hyper-Variable Number of Tandem Repeats in Intron 8 of the Human Dopamine Transporter Gene (*SLC6A3*)

David J. Vandenbergh[1,2,3,4], Abner T. Aplsey[1,2], Max A. Verbiest[5,6,7], Wesley J. Stone[1], Emma R. Domico[1,8], Carly A. Brogan[1], Evan R. Buck[1], Kathleen M. Cardone[1], Andrew J. Burich[9], Maria Anisimova[5,7]

[1]Department of Biobehavioral Health, [2]Molecular, Cellular, and Integrative Biosciences Program, [3]Institute of the Neurosciences, [4]Bioinformatics and Genomics Program, The Pennsylvania State University, [5]Institute of Computational Life Science, School of Life Sciences and Facility Management, Zurich University of Applied Sciences, [6]Department of Molecular Life Sciences, University of Zurich, [7]Swiss Institute of Bioinformatics, [8]Current: Armstrong Institute for Patient Safety and Quality, Johns Hopkins University, [9]Department of Information Science and Technologies, The Pennsylvania State University

Genetic analysis of *SLC6A3* has generated conflicting results for addiction phenotypes. This conflict might arise from an inability to test the appropriate loci in the gene. We identified multiple Tandem Repeats in *SLC6A3* with Tandem Repeat Annotation Library (TRAL), and performed initial characterization of those that varied in copy number (VNTRs) from 64 unrelated long-read haplotype-phased *SLC6A3* sequences. Sequence similarity of each repeat unit of the five VNTRs is reported, along with the correlations of SNP-SNP, SNP-VNTR and VNTR-VNTR alleles across the gene. One VNTR is hyper-variable (hyVNTR) and is located in intron 8 of *SLC6A3*. It contains a range of 3.4-133.4 repeat copies with 46 alleles in the 64 chromosomes. The consensus sequence of the repeat unit is 38 bp with 82% G+C content.

The repeat is predicted to form G-quadruplexes (G4s) *in silico*, which was confirmed by circular dichroism spectroscopy. This hyVNTR contains multiple putative binding sites for PRDM9 (recruiter of recombination promoting proteins), and it is in weak linkage disequilibrium with flanking genetic markers, suggesting it is a hotspot for recombination, and that previous studies could not estimate genetic effects of this hyVNTR.