**SUPPLEMENTARY MATERIAL**

**Enrichment of Motilin Receptor loss of function variants in Gastroparesis**

Sandra P. Smieszek, PhD, Jesse L. Carlin, PhD, Changfu Xiao, PhD, Gunther Birznieks, MS, Christos M. Polymeropoulos, MD, Mihael H. Polymeropoulos, MD

1Vanda Pharmaceuticals Inc., 2200 Pennsylvania NW, Suite 300-E, Washington, DC 20037

**corresponding author: sandra.smieszek@vandapharma.com**

**METHODS**

**Datasets and Statistics**

The first gastroparesis dataset (phase II clinical study - VP-VLY-686-2301) was a multiethnic cohort (n -119). This cohort was composed of 90.7% females; the average age was 45.9; 86.2% of the patients self-identified as White, 10.5% self-reported as Black or African American,  2% self-reported American Indian or Alaska Native and 1.3% reported as Asian. The replication dataset consisted of WGS samples collected as part of an ongoing gastroparesis study (phase III gastroparesis clinical study (VP-VLY-686-3301)). This was a multiethnic cohort composed of 70.2% females; the average age was 48.5; 64.5% of the patients self-identified as White, 31.7% self-reported as Black or African American,  0.8% self-reported American Indian or Alaska Native and 1.9% reported as Asian.

For comparisons of variant frequencies, we used both gastroparesis cohorts as well as internal controls. All cases and controls were processed with the same bioinformatic pipeline for variant calling. We also compared allelic frequencies to those reported in gnomAD6.  For ancestry definition for the African American specific analysis, we performed Principal Component Analysis on a set of predefined variants prior to doing allelic frequencies counts.

For the pLOF variants, we used basic quality control filters (only variants passing GATK’s Variant Quality Score Recalibration - VQSR) internal MAF of < 0.05 and restricted the variant effect to frameshift, stop gained, splice acceptor, and splice donor annotated by Annovar. In a follow-up analysis, we included all the missense variants in the motilin receptor in addition to the pLOF variants while keeping all other filters the same. Odds ratio and relative risk calculations were made on allelic frequencies of the identified variants of interest. For allelic enrichment, we used Fisher’s exact test. We report uncorrected P values, ORs, and 95% CIs. Since the main replication attempt involved a single test, we considered P values below 0.05 to be statistically significant.

Motilin Receptor protein structure was obtained from Alphafold9.

### DNA extraction library prep and sequencing

Incoming [nucleic acid](https://www.sciencedirect.com/topics/medicine-and-dentistry/nucleic-acid) samples are quantified using fluorescent-based assays (PicoGreen) to accurately determine whether sufficient material is available for library preparation and sequencing. DNA sample size distributions are profiled by a Fragment Analyzer (Advanced Analytics) or BioAnalyzer (Agilent Technologies), to assess sample quality and integrity. HumanCoreExome 24v1.3 array was performed on all human DNA samples sequenced. [Whole genome sequencing](https://www.sciencedirect.com/topics/medicine-and-dentistry/whole-genome-sequencing) (WGS) libraries were prepared using the Truseq DNA PCR-free Library Preparation Kit. Whole Genome data were processed on NYGC automated pipeline. Paired-end 150 bp reads were aligned to the GRCh37 human reference (BWA-MEM v0.7.8) and processed with GATK best-practices workflow (GATK v3.4.0). The mean coverage was 35.8, it reflects the samples average. All high quality variants obtained from GATK were annotated for functional effects (intronic, intergenic, splicing, nonsynonymous, stopgain and frameshifts) based on RefSeq transcripts using Annovar31. Additionally, Annovar was used to match general population frequencies from public databases (Exac, gnomAD, ESP6500, 1000 g) and to prioritize rare, loss-of-function variants.