The copy number variations at genes related to neuronal functions under selection in great tit

Vinicius H. da Silva[1,2,3], Veronika N. Laine[3], Mirte Bosse[3], Kees van Oers[3], Phillip Gienapp[3], Marcel E. Visser[1,3], Richard P.M.A Crooijmans[1], Martien A.M. Groenen[1]

Affiliation(s): Animal Breeding and Genomics Centre, Wageningen University[1], Swedish University of Agricultural Sciences[2], Netherlands Institute of Ecology (NIOO-KNAW)[3].

The great tit (*Parus major*) is a well-studied wild bird which has been used as a model species to document the effects of global warming on nature. The recent completion of a reference genome sequence of the great tit and the availability of a high density (HD) 500K SNP-chip, have enabled detailed genomic studies in this species. These genomic tools allowed precise and scalable measures of genetic and non-genetic (i.e. epigenetic marks) variation, identification of SNP variants under selection and copy number variations (CNV - **Figure 1**).
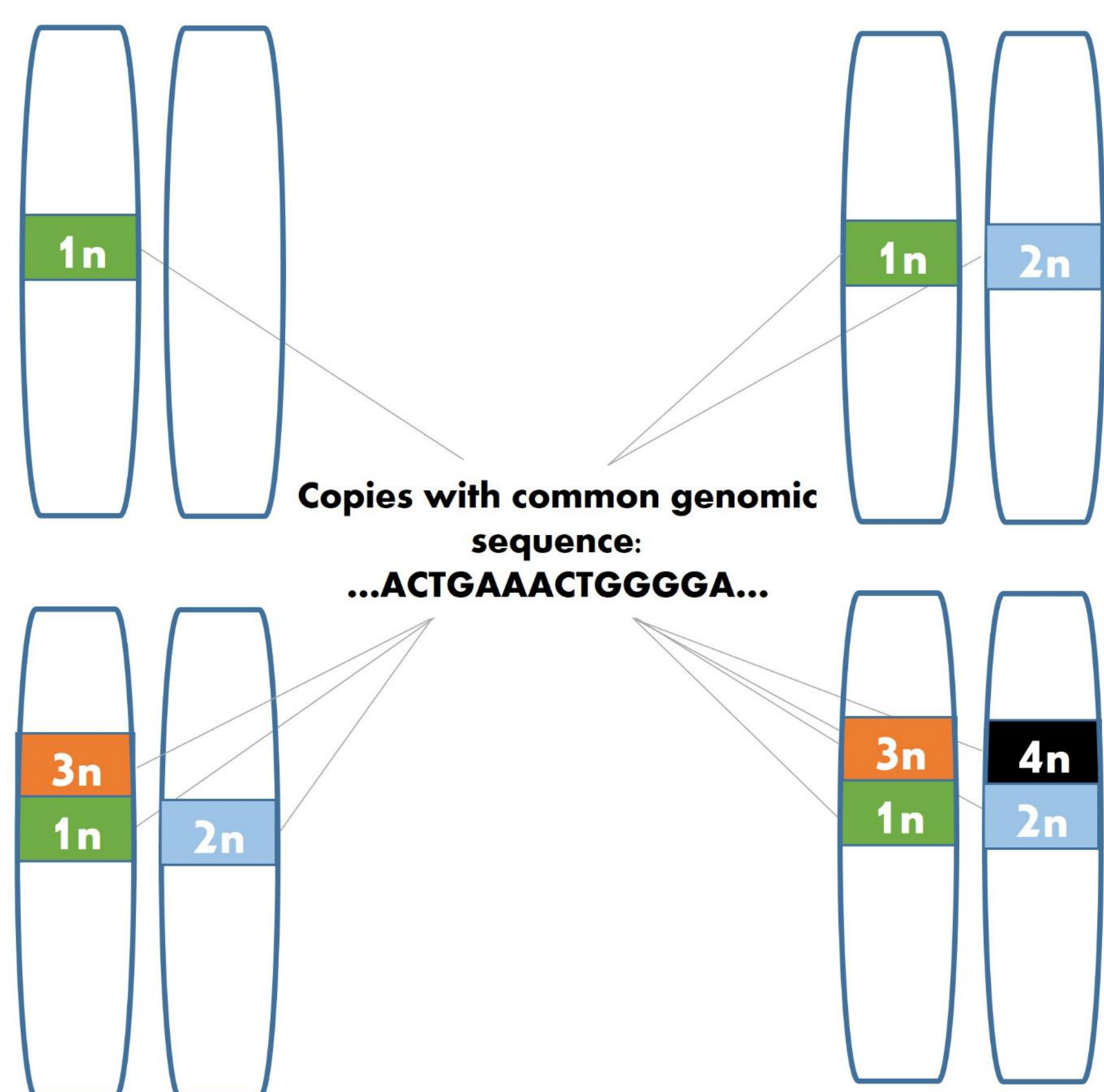

Figure 1. **Copy number state concept ranging from one copy (1n) to four copies (4n) in a specific genomic interval for a given animal.** SNP-arrays can infer with reasonable quality states ranging from 0n to 4n.

Genes related to neuronal functions were previously identified at regions under positive selection and prone to be methylated in great tits. This suggests that complex variation (CNV + SNP + methylation) plays an important role in great tit microevolution. A representative example is the gene *CD200* (**Figure 3**) which is associated with Parkinson´s disease, neuroimmunity, and is under positive selection in the Dutch great tit population and overlaps a highly frequent CNVR.

Figure 3. **Duplication CNVs at *CD200* genomic region (chromosome 1).** The chromosome ideogram depicts the CNVR position in red. The first track contain the CN frequencies (% of animals with 3n state). The 'G' track indicates the genomic position of the *CD200* gene. Each of the remaining tracks show one random animal with it respective CNV boundaries.

An initial CNV mapping in great tit genome, using the HD SNP-chip intensities and allele frequencies, from over 2000 females, is depicted in **Figure 2**.
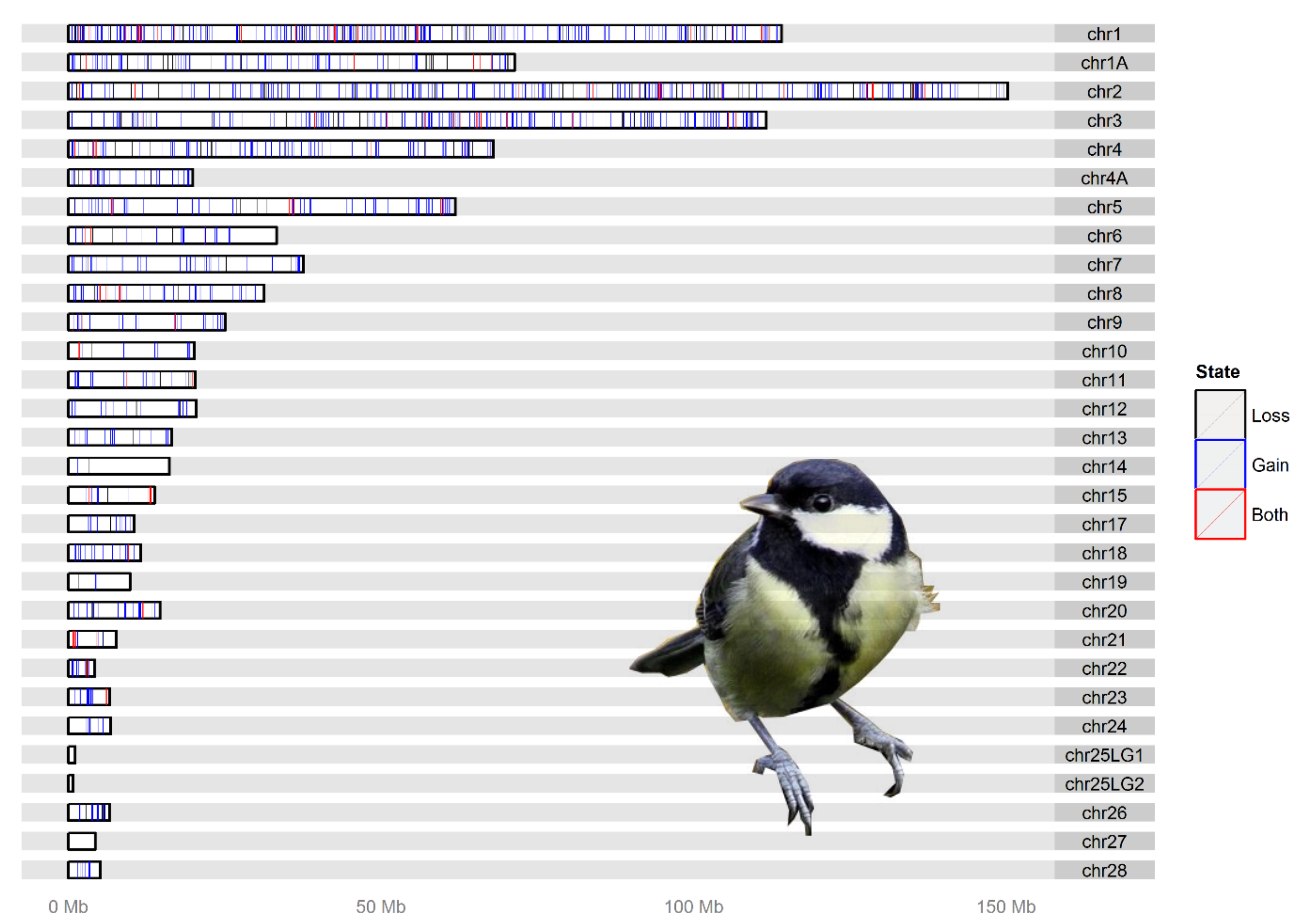

Figure 2. **Chromosomal distribution of regions of CNV (CNVRs) in the great tit population.** CNVRs with 0n and/or 1n were considered 'Loss', with 3n and/or 4n considered 'Gain' and overlapping 'Gain' and 'Loss' CNVRs were merged into 'Both' regions.

We found regions of CNVs (CNVRs) significantly enriched for neuronal related pathways as '*neuroactive ligand-receptor interaction*' (**Figure 4**) and '*Gap junction*'.
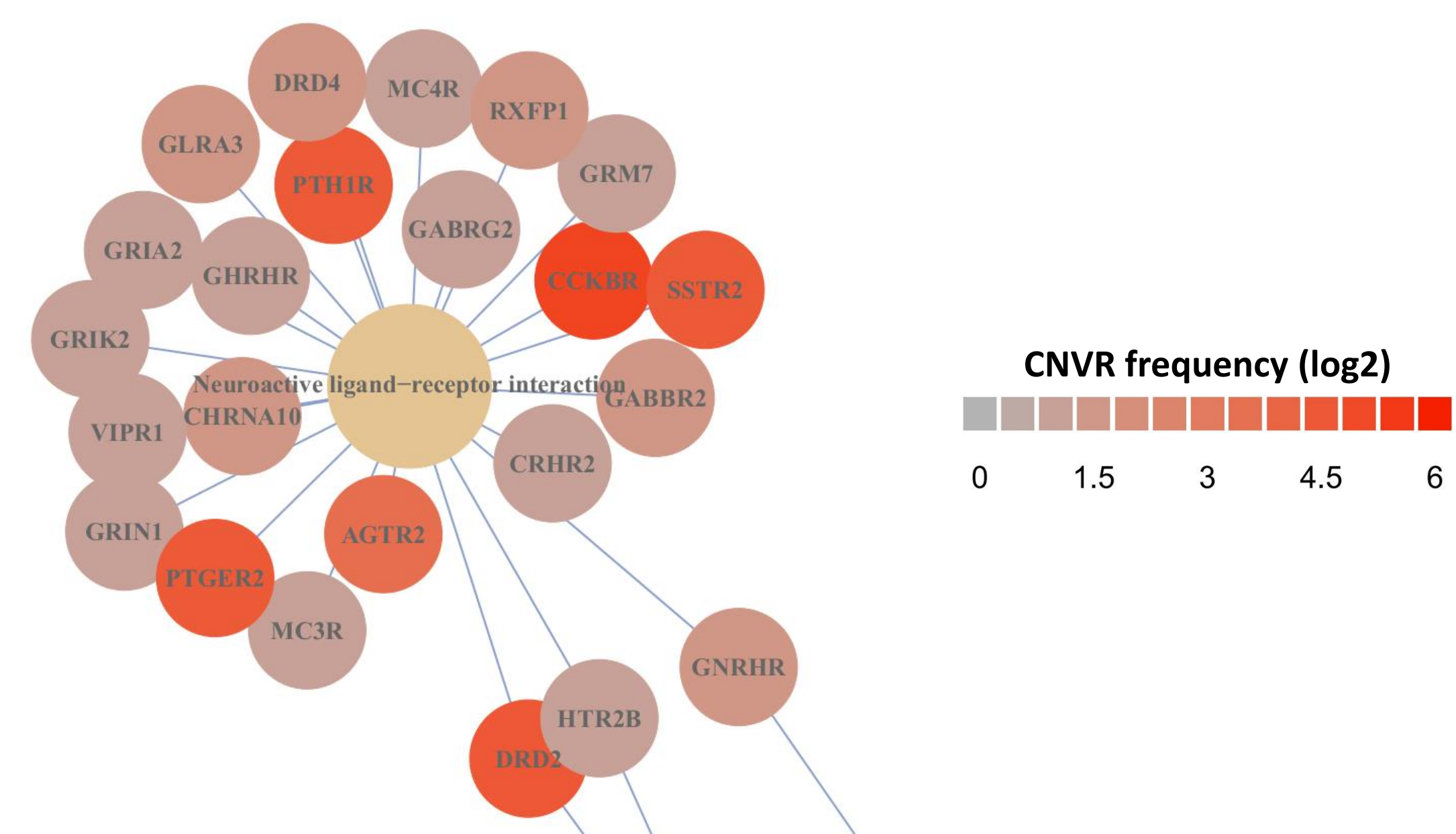

Figure 4. **Neuronal related KEGG pathway and it respective CNV frequencies for each gene in a population of >2000 females.** The first track show the chromosome coordinates, the second track shows the frequency of 3n state in this genomic region and each of the remaining tracks represent one random animal with CNV.

The CNV analysis is at an early stage and the next steps include strategies to improve the SNP-chip overall signal quality, SNP-CNV haplotype reconstruction, a selection analysis as well as a genome-wide association study for seasonal phenotypes (e.g. timing of breeding) changing as a result of global warming.