

## ABSTRACT

Schizophrenia (SCZ) and Bipolar disorder (BP) are severe neuropsychiatric disorders with an estimated prevalence of 1% and 0.8%, respectively. Both disorders have outstanding heritabilities, estimated between 60% and 80%. Recent genome-wide association studies (GWAS) have just started to shed light on the genetic architecture of these complex traits. According to these studies i) SCZ and BP are highly polygenic disorders, ii) thousands of genetic loci contribute to the disease risk and iii) common variation explains an important proportion of these complex traits. Polygenic risk scores, (PRS), summarize the joint risk effect of such common risk variants. However little is known about the biological processes, cellular pathways and/or cell types underlying such a polygenic risk. Therefore, we present an integrated computational approach combining transcriptomic and proteomic data at cellular resolution with GWAS data sets from the KFO241 consortium with the aim to identify a possible cell-selective genetic disease burden in individual patients. This approach towards patient classification may allow us to define working hypothesis that may be experimentally addressed in the future in corresponding patient-derived cellular systems.

## INTRODUCTION

Schizophrenia (SCZ) and Bipolar Disorder (BD) are severe neuropsychiatric disorders with high heritabilities, estimated between 60% and 80%. Recent genome-wide association studies (GWAS) have just started to shed light on the genetic architecture of these complex traits (Ripke et al., 2014; Charney et al., 2013). According to these studies i) SCZ and BD are highly polygenic disorders, ii) thousands of genetic loci contribute to the disease risk and iii) common variation explains an important proportion of these complex traits (Gratten et al., 2014). Polygenic risk scores (PRS) summarize the joint risk effect of such common risk variants. However little is known about the biological processes, cellular pathways and/or cell types underlying such a polygenic risk. We hypothesise that at least some groups of patients will show profiles of genetic risk with a high specificity regarding the cell types involved. Therefore, the objective of this study was to generate cell type-specific PRS in a sample of SCZ and BD patients.

## METHODS

### Subjects

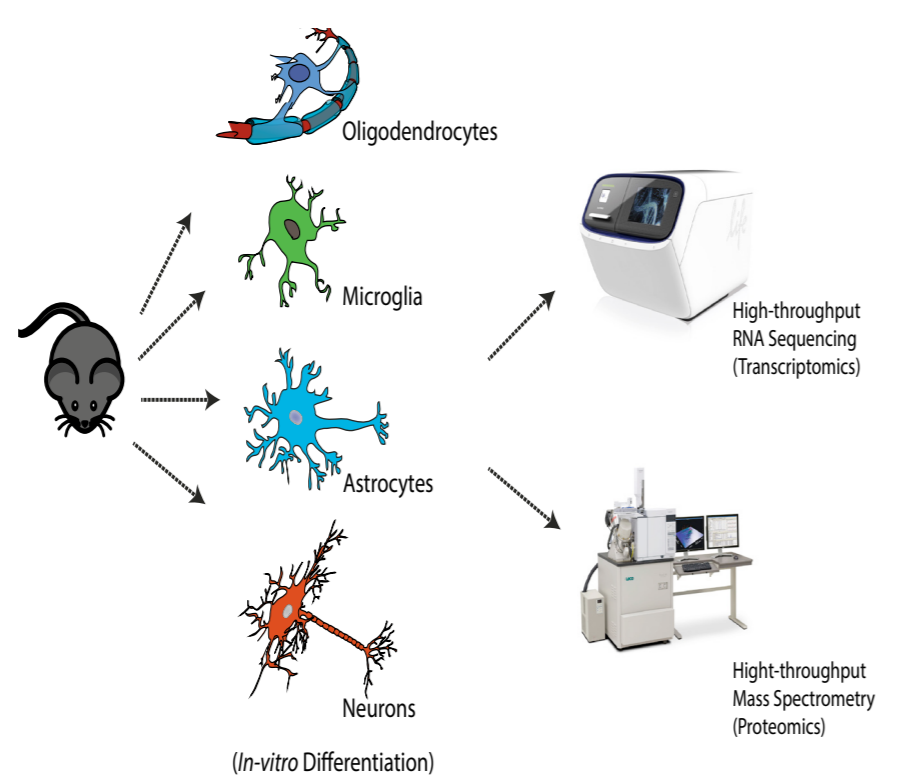
The sample under analysis consisted of 390 SCZ (or schizoaffective) and 270 BD patients belonging to the KFO241 cohort (www.kfo241.de). These patients were diagnosed according to DSM-IV-TR criteria. A sample of 503 EUR subjects from the general population belonging to the 1000 Genomes Project collection was also analysed. The patients samples were genotyped using the Infinium PsychArray Bead-Chip (Illumina®).

### Schizophrenia PRS calculation

For the calculation of the different PRS, SNPs were selected using the latest SCZ GWAS (Ripke et al., 2014) as initial training sample. This information was used to construct a score in our independent replication sample by forming the weighted sum of associated alleles within each subject across different P-value thresholds. PRS for each cell type considered only those SNPs mapping to the genomic coordinates (hg19) from those genes defined in cell type-specific lists.

### Expression profiles for major cell types in CNS & definition of cell type-specific gene-sets

Primary mouse brain cells from new born mice were differentiated *in vitro* into Astrocytes, Microglia, Neurons and Oligodendrocytes. The cells were then lysed at different stages of development/maturation for parallel high-throughput RNA-seq & mass spectrometry.



These experiments yielded lists of cell type-specific genes at the transcriptomic and proteomic levels in mice. The human homolog genes were accordingly matched in order to generate the human cell type-specific gene-sets for PRS calculations.

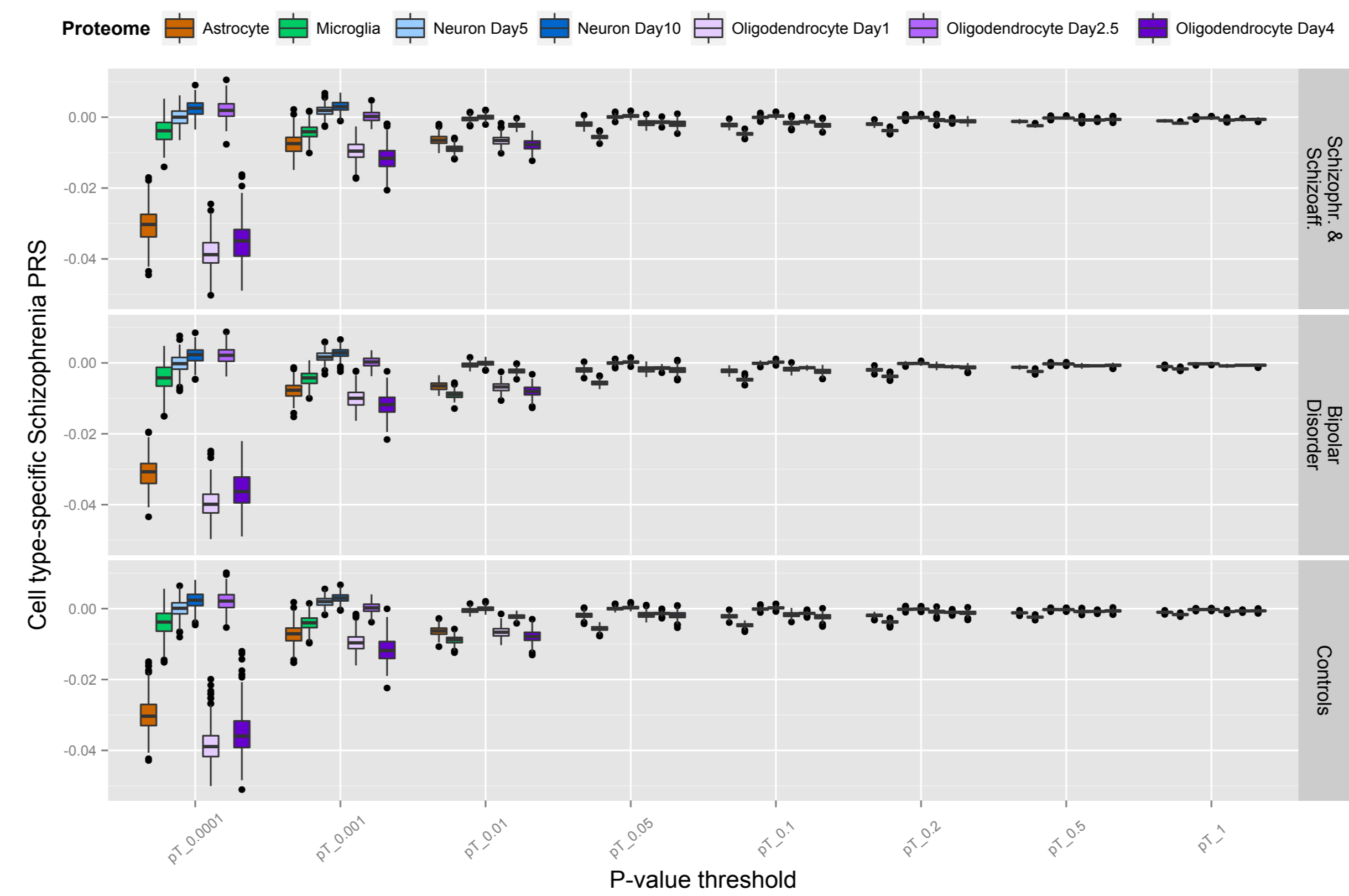
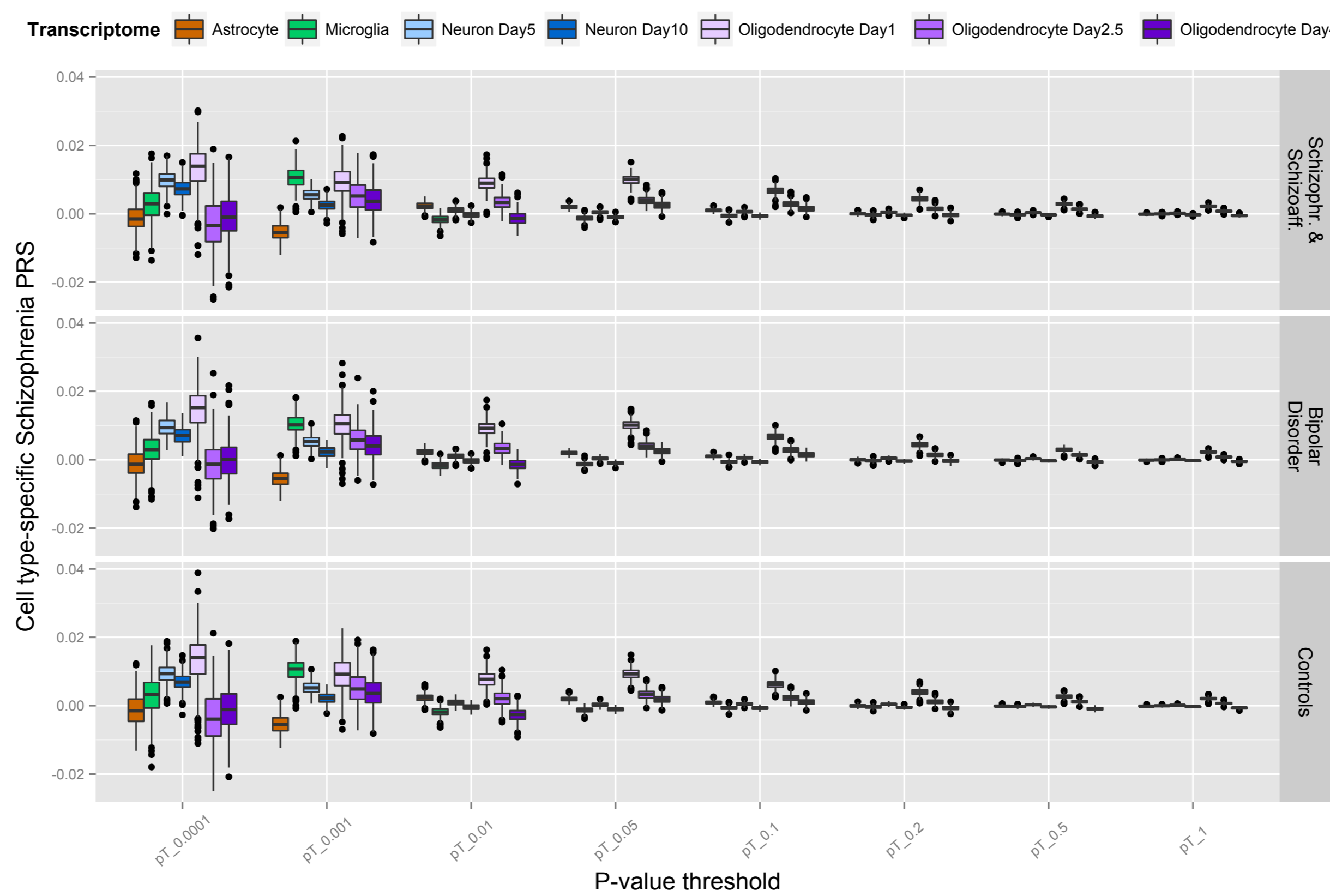
### Statistical analyses

PLINK 1.07 (Purcell et al., 2007) and R were used for data manipulation. PRS were calculated with PRSice (Euesden et al., 2015) based on 317,071 SNPs available both in the patients and general population datasets. SPSS v22 was used for comparing PRS between groups.

## RESULTS

### Figure 1. Patterns of cell type-specific PRS according to transcriptomics and proteomics

SCZ, BD patients and the control EUR sample showed extremely similar patterns in the cell type-specific PRS across the different P-value thresholds. None of the cell type scores showed a remarkable increase/decrease comparing SCZ, BD patients and control EUR sample ( $P > 0.05$  for all comparisons). This was observed both at using transcriptome-derived (Upper panel) and proteome-derived (Lower panel) gene-sets.

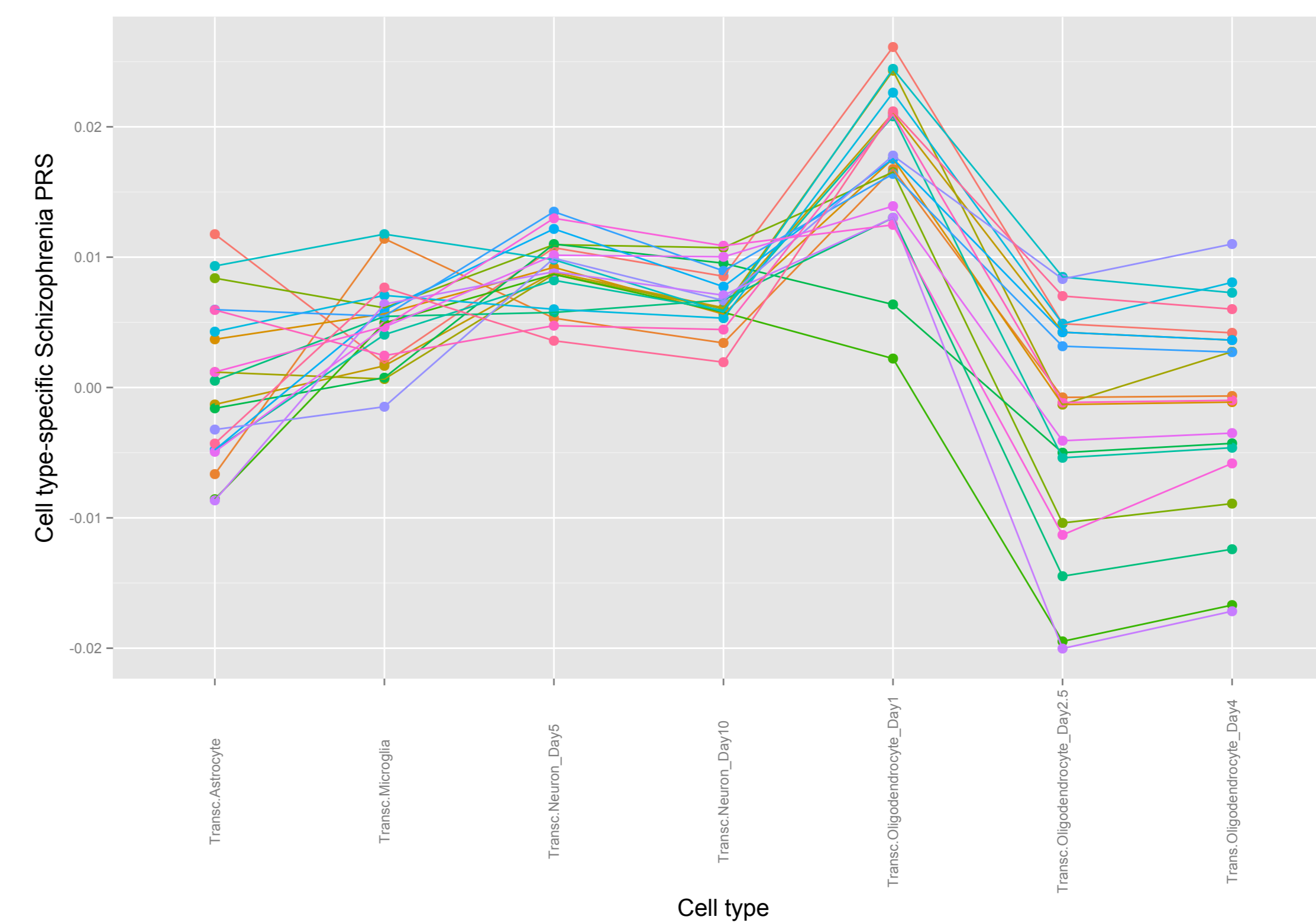


## REFERENCES

- Charney et al. (2013), XXI World Congress Psychiatric Genetics (Boston).  
 Euesden et al. (2015) Bioinformatics 31(9):1466-8.  
 Gratten et al. (2014) Nature Neuroscience 17(6):782-90.  
 Purcell et al., (2007) Am J Hum Genet 81(3):559-75  
 Ripke et al. (2014) Nature 511(7510):421-7.

### Figure 2. Individual cell type-specific risk profile

At the individual level, a differential polygenic load was observed across the different cell type-specific polygenic risk profiles. As an example, a random set of 20 patients (each one represented by a color) and their transcriptome-based cell type-specific polygenic risk scores at P-value threshold=0.0001 are displayed in this figure. The genetic burden estimated with such cell type-specific PRS shows that not all patients carry the same genetic load with respect to the different cell types. Very similar results are obtained using proteomic gene-sets and at different P-value thresholds.



## DISCUSSION

### Results at the group level:

Polygenic risk scores based on the gene-sets defined by transcriptomics or proteomics do not seem to differ between cases (SCZ, BD) and healthy controls. Likewise the same polygenic scores can not discriminate between SCZ and BD patients.

### Results the individual level

Each patient shows a differential profile of polygenic scores with respect to the different cell types, both for those transcriptomic-derived and proteomic-derived.

Taken together, the results of this study suggest that cell type-specific genetic factors may be useful for distinguish subgroups of patients. The validity of such subgroups still needs to be ascertained at the biological, phenotypical and clinical levels.

## ONGOING WORK...

Genotype Imputation currently ongoing using 1000 genomes Phase3 release as reference dataset. Number of SNPs available will grow from ~500,000 to ~ 7-9M). This may increase polygenic scoring accuracy.

Statistical validation of the findings at the individual level: in order to discard random effects on the different profiles within each patient.

## OUTLOOK

According to the individual profile, candidate patients may be recruited for the generation of induced pluripotent stem cells (iPSCs), subsequently differentiated into the cell type of interest. This approach towards patient classification may allow us to define working hypothesis that may be experimentally addressed in the future in corresponding patient-derived cellular systems.

## ACKNOWLEDGEMENTS / FINANCIAL DISCLOSURE

We thank all of the patients for participating in this study. This research was funded by the Deutsche Forschungsgemeinschaft (DFG) grants: Klinische Forschergruppe (KFO) 241: TP1 (SCHU 1603/5-1), FKZ RO4076/1-1 and FKZ RO4076/3-1. Thomas G. Schulze was supported by the Lisa-Oehler-Foundation.

## CONFLICT OF INTEREST

There are no conflicts of interest.