**Principal Isoforms for Selected Genes**

(ones in green are completed)

EEF1A2 - **ENST00000217182.6**, **NM\_001958.5;** corresponds to CCDS13522, NM\_001958.5, NP\_001949.1. (good expression, mostly in brain and heart) – used in literature (same in GRCh37; the same one used for the alignment in DIOPT)

STXBP1 - **ENST00000373299.5**, **NM\_001032221.6;** corresponds toCCDS35146, NM\_001032221.6, NP\_001027392.1 (good expression across tissues) (same in GRCh37; DIOPT uses different protein).

PTPN11 - ENST00000351677.7 in GRCh38 (ENST00000351677.2 in GRCh37), corresponds to NM\_002834.5, NP\_002825.3, CCDS9163. DIOPT uses different protein isoform.

~~TCF4~~ - ENST00000354452.8, NM\_001083962.2 (no expression in GTEx; highly expressed ENST00000635990.2 which is not a protein-coding transcript in Ensambl). A recent paper (PMID: 34837432) on clinical interpretation of variants in TCF4 (among few other genes) has chosen the above isoform as the one relevant for clinical interpretation.

MECP2 - ENST00000453960.7, NM\_001110792.2 (**E1 isoform**) (no brain expression in GTEx, but this is probably not correct) **OR** **ENST00000303391.11**, **NM\_004992.4** (**E2 isoform**) (brain expression; used for clinical interpretation PMID: 34837432). The Ensemble transcript IDs are the same across assemblies.

CTNNB1 - **ENST00000349496.11**, **NM\_001904.**4 (good expression across tissues). The ID is the same for both GRCh37 and GRCh38

DDX3X - ENST00000644876.2, NM\_001356.5 (in GRCh38), **ENST00000399959.2**, (**NM\_001356.5, NP\_001347**) in GRCh37. It is not found in GTEx but used in ClinVar, VariCarta and in the literature. A paper with biochemical variants (PMID: 32135084) added to Zotero.

~~PTPN1 -~~ **~~ENST00000371621.5~~**~~,~~ **~~NM\_002827.4~~** ~~(expression in GTEx, used in ClinVar, VariCarta; literature focused on common variants for GWAS in diabetes) (same in GRCh37; the same one used for the alignment in DIOPT)~~ – WRONG GENE

BRAF - ENST00000644969.2, NM\_001374258.1. ClinVar mostly uses **NM\_004333.6 (NP\_004324.2)**; the same with VariCarta; this is also the most prevalent isoform in the literature. This corresponds to Ensembl’s transcript **ENST00000288602.6** (in GRCh37; ENST00000646891.1in GRCh38). Ensembl changed the name of the transcript but the sequence remained the same (both IDs are linked to CCDS5863). The Ensembl and NCBI (NM\_) ID are not directly linked in GRCh38 but are linked in GRCh37.

