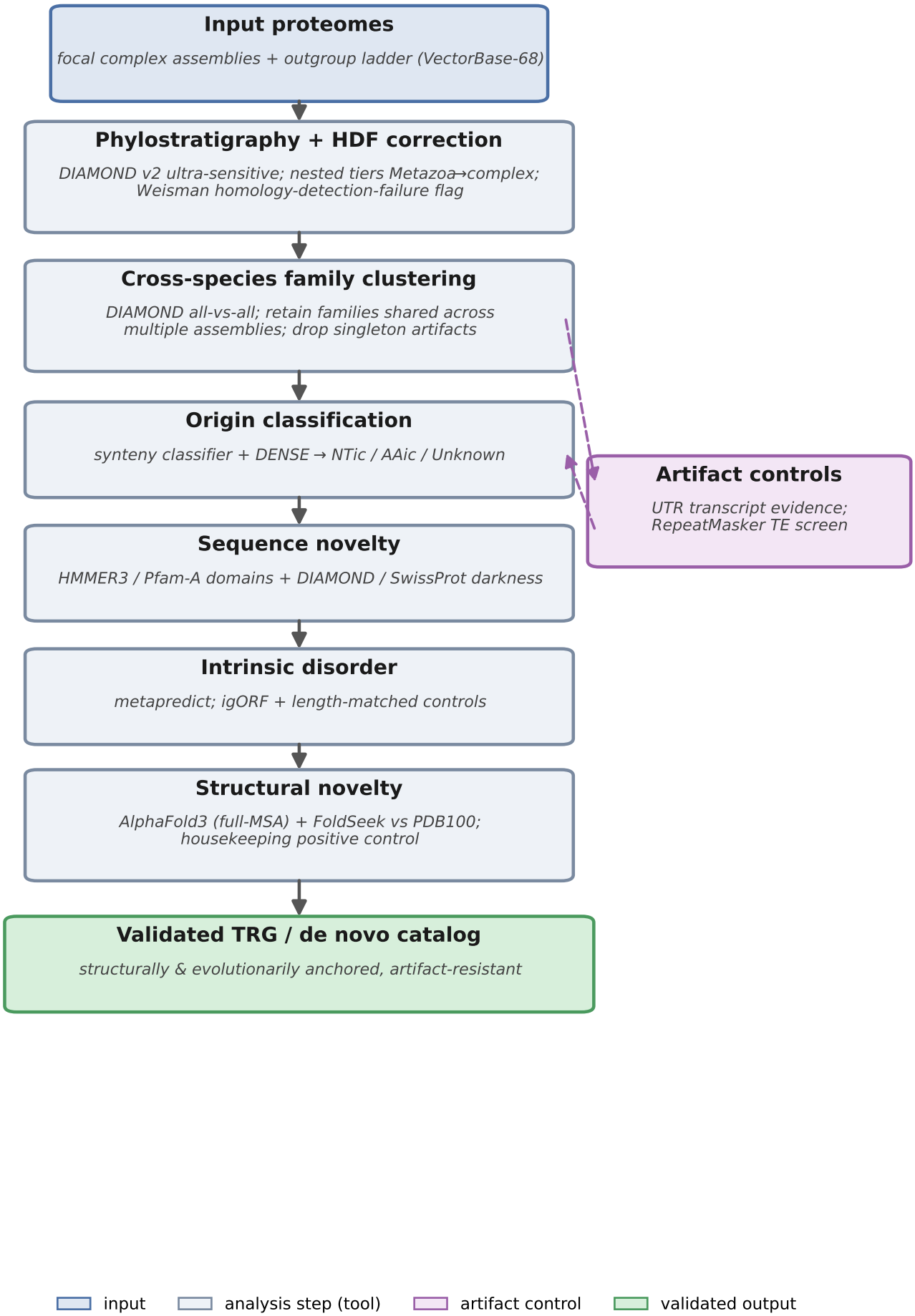Repeatable bioinformatic pipeline for TRG / de novo gene discovery and validation