

## Streszczenie

Od odkrycia u ssaków pierwszych nakładających się genów kodujących białka minęło ponad trzydzieści lat. Przez długi czas fenomen ten uważany był za marginalny, jednakże z biegiem lat przybywało przykładów par genów nakładających się u wielu gatunków. Zjawisku nakładania się genów przypisano wiele funkcji regulatorowych zarówno na poziomie transkrypcji, jak również post-transkrypcyjnie. Niemniej jednak po dziś dzień nie ustalono jednoznacznie wpływu zjawiska nakładania się genów na ich poziom ekspresji.

Wykorzystując koordynaty alternatywnych miejsc startu transkrypcji (TSS) w 73 ludzkich i 10 mysich bibliotekach, zidentyfikowano 582 ludzkie i 113 mysich par genów kodujących białka, nakładających się końcami 5'. Wykazano, że para genów wykorzystująca do ekspresji nakładające się miejsca TSS w jednej bibliotece, może w innej wykorzystywać alternatywne, nienakładające miejsca startu transkrypcji. Tkankowo specyficzne nakładanie się genów może być częściowo związane z rodzajem tkanki w których geny ulegają ekspresji. Wykazano jednak, że również warunki egzogenne, jak np. transfekcja, mogą mieć wpływ na zachodzenie nakładania, co ma związek z alternatywnym wykorzystaniem promotorów. Badania wykazały również, że przynajmniej część genów nakładających się jest regulowana przez promotory dwukierunkowe. Wykazano ponadto, że poziom ekspresji genów nakładających się jest przeciętnie wyższy aniżeli innych genów. Co więcej, w pracy zademonstrowano, że przy wykorzystaniu nakładających się miejsc TSS, geny nakładające się mają wyższy poziom ekspresji aniżeli te same geny, gdy ulegają one ekspresji z nienakładających się miejsc startu transkrypcji.

Nakładanie się genów kodujących białka zostało również przestudiowane w kontekście stanu chromatyny i aktywności polimerazy RNA II. Analiza ta pozwoliła na zidentyfikowanie wzorców aktywności polimerazy oraz wykazanie, że zjawisko interferencji transkrypcji może zachodzić zarówno gdy geny wykorzystują nakładające się jak i nienakładające się promotory. W niektórych przypadkach brak podwyższonej aktywności polimerazy RNA II, a co za tym idzie brak interferencji w rejonie nakładania się genów, połączony został z monoalleliczną ekspresją genów nakładających się.

Kluczowe wyniki zaprezentowane w niniejszej pracy doktorskiej zostały zdeponowane w publicznie dostępnej bazie danych OverGeneDB, która znajduje się pod adresem internetowym <http://overgenedb.amu.edu.pl>.

## **Abstract**

Over 30 years have passed since the discovery of the first mammalian overlapping protein-coding genes. For a long time, gene overlap was considered to be rather uncommon, but nowadays more and more of different types of overlapping genes, depending on their position and the transcription direction, is reported in diverse species. Gene overlap have been shown to play various regulatory functions on transcriptional and post-transcriptional levels. However, the true influence of the gene overlap phenomenon on the overlapping genes' expression levels is still a matter of debate.

Here, using information of the alternative transcription start sites (TSS) in 73 human and 10 mouse libraries, a total of 582 human and 113 mice 5' end protein-coding overlapping gene pairs were identified. It was shown, that a single gene pair, identified as overlapping in one library, may use a different set of alternative TSS in different libraries, often resulting in the transcription from non-overlapping regions. Tissue-specific gene overlap patterns may partially be related with a tissue type, but it was also shown that gene overlap may be triggered by the environmental changes like the cells transfection, most possibly as a side effect of the alternative promoters' usage. Conducted studies revealed that at least some of the 5' end protein-coding overlapping genes may be regulated by the bidirectional promoters. It was also shown, that overlapping genes on average tend to have a higher expression level than non-overlapping genes. What is more, it was demonstrated that the expression level of overlapping genes is on average higher when they utilize overlapping promoters, than when the same genes are expressed from non-overlapping transcription start sites.

Protein-coding gene overlap was also studied in the context of the chromatin state and RNA polymerase II (RNAPII) activity. Although it was hard to identify transcriptional interference (TI) down-regulatory effect on the expression level based solely on the studies of the overlapping genes' expression levels, it was possible to find conceivable marks of TI within the RNAPII activity patterns. Interestingly these patterns were identified both when genes were overlapping and when they utilized non-overlapping promoters. In some cases, no enhanced RNA polymerase II activity was found within the overlap regions, which was suggested to be possibly connected with the overlapping genes monoallelic expression.

Finally, OverGeneDB database was created so that anyone interested in deeper understanding of the gene overlap phenomenon could explore the key results presented in this project. OverGeneDB is publically accessible under <http://overgenedb.amu.edu.pl>.