

Biografia naukowa

1. **Imię i nazwisko: Karol Kozak**

2. **Posiadane dyplomy, stopnie naukowe – z podaniem nazwy, miejsca i roku uzyskania oraz tytułu rozprawy doktorskiej:**

- magister nauk technicznych w zakresie Elektrotechniki i Informatyki, Politechnika Radomska (Polska) oraz Techniczny Uniwersytet w Dreźnie (Niemcy), 2001
- doktor nauk technicznych w zakresie Bioinformatyki, Wydział Informatyki Politechniki Śląskiej w Gliwicach oraz Instytut Maxa Plancka w Dreźnie (Niemcy), 2007
rozprawa doktorska: „Klasyfikacja danych z biologicznych analiz wieloprzepustowych za pomocą adaptacyjnych metod k-najbliższych sąsiadów” grudzień 2007, Wydział Informatyki Politechniki Śląskiej w Gliwicach (promotor: prof. dr hab. Katarzyna Stąpor).

3. **Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych:**

- 2003- 2007 Instytut Maxa Plancka (Monachium i Drezno, Niemcy)
- 2007- 2012 Uniwersytet, ETH Zurych (Zurych, Szwajcaria)
- 2012 - dziś:
 - Zatrudniony w Instytucie COMEDD, Fraunhofer oraz na Uniwersytecie Drezdeńskim, Wydział Medycyny, obecnie na stanowisku dyrektora działu informatyki i analizy danych bio-medycznych (Juniorprofesor). Dodatkowo jest zatrudniony jako gościnnie profesor i pracownik naukowy na:
 - LMSC, ETH Zurych, Szwajcaria jako gościnnie pracownik naukowy
 - Indian Association for the Cultivation of Science jako wizytujący junior professor

4. **Wskazanie osiągnięcia wynikającego z artykułu 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.):**

4.1 Osiągnięcia naukowe

Cykl publikacji z lat 2007-2013, pod wspólnym tytułem „Bioinformatyka interferencji RNA – wpływ jakości modelowania cząsteczek RNAi na przewidywanie nowych funkcji genów w komórkach ssaków”.

4.2 **Lista publikacji stanowiących osiągnięcie naukowe (autor/autorzy, tytuł/tytuły publikacji, rok wydania, nazwa wydawnictwa)**

Blawert

1. Shaoli D., Ghosal S., Kozak K. and Chakrabarti J.: Customized siDesigner: A tool for designing siRNAs with functional off-target filtering. Journal of Biomolecular Structure & Dynamics, 2012
Impact Factor (IF): 4.98 z roku 2011(w roku 2011 nastąpiła zmiana indeksu, w roku 2012 będą ogłoszone nowe statystyki), **MNiSW:** -
wkład własny (K. Kozak, wkład procentowy 40%):
wspólne prowadzenie zespołu, zebranie danych eksperymentalnych do badań we współpracy z laboratoriami w Szwajcarii; przeważający udział w analizie skutków ubocznych siRNA i interpretacji rezultatów badań oraz w przygotowaniu publikacji (w tym ujęcie i dyskusja problemu i poszczególnych zagadnień, koncepcja bioinformatyczna, filtrowanie skutków ubocznych), wewnętrzna recenzja pracy.
wkład współautorów publikacji (w postaci określenia słownego podany jest w ich oświadczeniach w Załączniku.
2. Kozak K.: The Use of Design Specificity in Standardized Mean Difference for Analysis of High throughput RNA Interference Screens. Lecture Notes in Computer Science 7451 Springer 2012, ISBN 978-3-642-32394
IF: - **MNiSW:** - **Book Chapter**
wkład własny (K. Kozak, wkład procentowy 100%):
pełny wkład własny: pomysł, koncepcja, analiza danych, organizacja danych eksperymentalnych, przygotowanie manuskryptu, korekta pracy po recenzjach.
3. Kozak K., Csucs G.: Kernelized Z' Factor in Multiparametric Screening Technology. RNA Biology, 2010, Vol 7, Issue 5, doi: 10.4161/rna.7.5.13239
IF: 5.59 **MNiSW:** 35
wkład własny (K. Kozak, wkład procentowy 80%):
pomysł i koncepcja badań i pracy, prowadzenie symulacji i analiz komputerowych; główny udział w analizie i interpretacji rezultatów badań oraz w przygotowaniu publikacji (przygotowanie manuskryptu, obrazów, wykresów), korekta pracy po recenzjach.
wkład współautorów publikacji (w postaci określenia słownego podany jest w ich oświadczeniach w Załączniku #.
4. Mazur S., Csucs G., Kozak K.: Z' Factor including siRNA design quality parameter in RNAi screening experiments, RNA Biology, 2012
IF: 4.93, **MNiSW:** 35
wkład własny (K. Kozak, wkład procentowy 70%):
pomysł i koncepcja badań i pracy, zebranie danych eksperymentalnych do badań we współpracy z laboratoriami w Szwajcarii, prowadzenie projektu programistycznego; przeważający udział w analizie i interpretacji rezultatów badań oraz w przygotowaniu publikacji (w tym ujęcie i dyskusja problemu i poszczególnych zagadnień, koncepcja bioinformatyczna, modelowanie bazy danych, wnioski), korekta pracy po recenzjach.
wkład współautorów publikacji (w postaci określenia słownego podany jest w ich oświadczeniach w Załączniku.
5. Zwolinski L, Kozak M, Kozak K.: 1Click1View: interactive visualization methodology for RNAi cell based microscopic screening. BioMed Research International, vol. 2013.
IF: 2.436 **MNiSW:** 20

wkład własny (K. Kozak, wkład procentowy 80%):

pomysł i koncepcja badań oraz pracy, prowadzenie symulacji i analiz komputerowych; główny udział w analizie i interpretacji rezultatów badań oraz w przygotowaniu publikacji (przygotowanie manuskryptu, obrazów, wykresów), korekta pracy po recenzjach.

wkład współautorów publikacji (w postaci określenia słownego podany jest w ich oświadczeniach w Załączniku.

6. Kozak K.: Annotation and specificity of existing genome-wide siRNA libraries. *Nucleic Acid Therapeutics*, 10.1089/nat.2012.0387, 12. 2012

IF 2.79 MNiSW: 25

wkład własny (K. Kozak, wkład procentowy 100%):

pomysł, koncepcja, analiza danych, organizacja danych eksperymentalnych, przygotowanie manuskryptu, korekta pracy po recenzjach.

7. Mazur S., Csucs G., Kozak K.: RNAiAtlas: a database for RNAi (siRNA) libraries and their specificity, *Oxford Bioinformatics*, 2012

IF: 3.179 MNiSW: -

wkład własny (K. Kozak, wkład procentowy 80%):

pomysł i koncepcja badań i pracy, zebranie materiału do badań w Szwajcarii, prowadzenie projektu programistycznego; przeważający udział w analizie i interpretacji rezultatów badań oraz w przygotowaniu publikacji (w tym ujęcie i dyskusja problemu i poszczególnych zagadnień, koncepcja bioinformatyczna, modelowanie bazy danych, wnioski), korekta pracy po recenzjach.

wkład współautorów publikacji (w postaci określenia słownego podany jest w ich oświadczeniach w Załączniku.

4.3 Streszczenie wyników uzyskanych w publikacjach stanowiących osiągnięcie naukowe (kolejność chronologiczna)

OBIEKT BADAŃ:

Odkrycie zjawiska interferencji RNA zapoczątkowało nowe badania dotyczące hamowania ekspresji wybranych genów przez małe interferujące RNA (RNAi: siRNA, esiRNA) w komórkach ssaków. Obiektem badań są cząsteczki RNAi i ich funkcje używane w wysoce sprawnych badaniach przesiewowych (WSBP, ang. HCS – high content screening).

CELE BADAŃ:

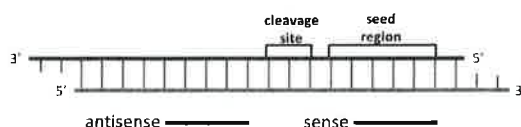
- wynalezienie nowych metod wyszukiwania docelowych związków (siRNA, esiRNA) wykorzystanych w WSBP
- wynalezienie nowych metod bioinformatycznych do modelowania cząsteczek RNAi
- wynalezienie nowych metod bioinformatycznych do oceny jakości RNAi
- badanie wpływu ewolucji anotacji genomu na jakość i modelowanie cząsteczek siRNA
- wynalezienie nowych metod bioinformatycznych do oceny jakości WSBP

OMÓWIENIE WYNIKÓW BADAŃ:

Moje osiągnięcia naukowe składają się z szeregu tematycznie połączonych publikacji dotyczących wpływu doboru cząsteczek RNAi na proces wyciszania genów podczas

eksperymentów WSBP. 20 grudnia 2002 roku czasopismo *Science* ogłosiło odkrycie procesów zjawiska interferencji RNA (RNAi), które było przełomowym odkryciem roku 2002. Odkrycie to stało się przyczyną do ożywienia w świecie biotechnologii, medycyny i farmacji, oraz zrewitalizowało dotychczas stosowane technologie terapeutyczne oparte na syntetycznych fragmentach kwasów nukleinowych. WSBP, w których używane są cząsteczki RNAi, pozwalają na wieloskalową identyfikację potencjalnych nowych funkcji genów. Proces wyboru aktywnych związków RNAi (faworytów) polega na ich wyodrębnieniu z całej kolekcji cząsteczek (kandydaci) branych pod uwagę. Poprzez aktywność cząsteczki rozumiemy jej oddziaływanie na komórkę według zasad ściśle określonych w specyfikacji analizy zwanej *protokołem*. Standardową techniką w odkrywaniu aktywnych cząsteczek jest ich ewaluacja za pomocą eksperymentu biologicznego. W roku 1990 zaczęto wprowadzać biologiczne *analizy* WSBP, które wykorzystują w pełni zautomatyzowane techniki potrzebne do przeprowadzenia biologicznego eksperymentu i pozwalające na przeanalizowanie tysięcy RNAi w bardzo krótkim czasie. W WSBP korzysta się zarówno z robotów, które automatyzują przepływ cieczy stosowanych w eksperymencie jak i z automatycznych mikroskopów fluorescencyjnych, które są w stanie w ciągu paru dni przeskanować i zgromadzić cyfrowe obrazy fluorescencyjne komórek. Niestety, (wciąż) wysokie koszty związane z przeprowadzeniem eksperymentów stanowią największą wadę WSBP.

Małe interferujące RNA (siRNA, small interfering RNA) to dwuniciowe cząsteczki RNA o długości ok. 20-25 par zasad (pz), które powodują wyciszenie ekspresji genów o homologicznej sekwencji (interferencja mRNA - RNAi). Charakteryzują się 100% homologią do sekwencji docelowej mRNA oraz w naturalnych warunkach odgrywają głównie rolę w obronie komórki przed wirusami. Powstają przez rozpięcie dwuniciowego RNA (np. wirusowego) w komórce przez enzym DICER na fragmenty odpowiedniej długości. siRNA wiążą się z kompleksem białkowym o aktywności rybonukleazy zwanym RISC. Kompleks ten wiąże się z cząsteczką mRNA komplementarną dla siRNA i rozcina ją, uniemożliwiając w ten sposób powstanie kodowanego przez nią białka. Zazwyczaj sekwencja siRNA jest w 100% komplementarna z sekwencją docelowego mRNA (Rysunek 1). Rośliny (i inne organizmy eukariotyczne) wykorzystują mechanizm interferencji RNA powodowany przez siRNA do obrony przed wirusami. Sztuczne siRNA powszechnie wykorzystywane są na szeroką skalę w biologii molekularnej i komórkowej oraz są obiektem intensywnych badań nad ich zastosowaniem w medycynie.



Rysunek 1: Struktura siRNA: 21 pz podwójnego RNA z przedłużonymi końcami 3', po każdej stronie (dwie zasady) 3'. Dwie nici są nazywane odpowiednio antysensową (lub aktywną lub nitką wprowadzającą) i sensową (lub nieaktywną), pierwsze 2-8 zasady nitki antisense są nazywane regionem zasiania (ang. seed region) i zasady 8-10 na nitce antisense nazywa się stroną dekoltu (ang. cleavage site).

Cząsteczki indukujące mechanizmy RNAi charakteryzują się specyficzną strukturą oraz specyficznymi właściwościami chemicznymi. Obie nici RNA o długości 21-23 nukleotydów, tworzą dupleks na odcinku 19 nukleotydów, pozostawiając na każdym końcu 3' po dwa lub

więcej niesparowane nukleotydy. Antysensowna nić siRNA tworzy komplementarny dupleks z docelowym mRNA. Modyfikacje w obrębie tej nici siRNA powodują obniżenie wydajności procesu wyciszania RNAi lub też całkowity zanik tego zjawiska. siRNA posiadają na końcach 3' wolną grupę hydroksylową, a na końcach 5' wolną grupę fosforanową.

Zaproponowano co najmniej dwa mechanizmy, według których może zachodzić proces wyciszania ekspresji genów. W pierwszym z nich inicjacja RNAi następuje przez konwersję dwuniciowego RNA do 21-23-nukleotydowych fragmentów za pomocą wielodomenowej rybonukleazy DICER, wywodzącej się z rodziny RNazy III. Produkty hydrolizy – krótkie dupleksy RNA, zwane siRNA, włączone są w kompleks nukleazowy RISC i kierują go do docelowej sekwencji mRNA. Endorybonukleaza obecna w kompleksie RISC wykorzystuje antysensowną nić siRNA do odnalezienia i degradacji komplementarnej sekwencji mRNA, dlatego też siRNA nazywany jest dupleksem kierującym. Nukleolityczne właściwości kompleksu RISC są istotą procesu interferencji za pomocą RNAi. W wielu firmach biotechnologicznych są prowadzone intensywne badania nad wprowadzaniem tych cząsteczek do leczenia a najbardziej zaawansowane badania znajdują się na II etapie badań klinicznych).

Cząsteczki siRNA mogą być tworzone chemicznie, enzymatycznie, lub generowane endogennie. Zasadniczą wadą każdej z tych technik jest ich koszt, jednak znaczącą zaletą – pełna kontrola sekwencji i struktury otrzymywanych cząsteczek. Najpowszechniejszą metodą otrzymywania cząsteczek siRNA jest wciąż ich chemiczna synteza. Wprowadzanie modyfikacji do syntetycznych siRNA pozwala nadać im cechy pożądane z punktu widzenia badań podstawowych jak i przyszłych zastosowań diagnostycznych. Wysoki koszt syntezy chemicznej RNA bez gwarancji, że wybrane sekwencje wywołają wydajne wyciszenie ekspresji docelowego genu, powoduje, że poszukuje się alternatywnych sposobów uzyskiwania i modelowania siRNA.

Jeden z problemów stanowi jednoznaczność wyciszania danego genu oraz skutki uboczne wywołane przez siRNA (ang. off-targets). Powodem tego jest nadal niewystarczająca wiedza na temat projektowania cząsteczek siRNA, które gwarantowałyby wysoką aktywności. Mechanizm off-target powoduje zmianę poziomu ekspresji wielu genów, co zostało zademonstrowane w szeregu badań. Procs projektowania jest bardzo istotnym elementem który wpływa na efekt syntezy i skutki off-target. W naszej dotychczasowej pracy (pozycja 1) moich osiągnięć naukowych *Shaoli D., Ghosal S., Kozak K. and Chakrabarti J.: Customized siDesigner: A tool for designing siRNAs with functional off-target filtering. Journal of Biomolecular Structure & Dynamics, 2012*) zademonstrowaliśmy nową metodę i narzędzie do modelowania cząsteczek siRNA pozwalają na redukcję skutków off-target. Modelowanie sekwencji siRNA, które mogą interferować z ekspresją wybranego genu wymaga wiedzy na temat sekwencji kodującego mRNA. W proponowanej metodzie wskazaliśmy że podczas modelowania bardzo istotne jest uwzględnienie większości istniejących transkryptów danego genu. Otrzymywane w ten sposób siRNA wywołują efekt wyciszenia genów egzo- i endogennych w różnych typach komórek ssaków. Stworzone przez nas oprogramowanie, SiDesigner, w połączeniu z dostępem do bazy sekwencji genomów różnych organizmów, wyszukuje potencjalnie aktywne sekwencje siRNA z uwzględnieniem redukcji skutków off-target.

W naszej publikacji zwróciliśmy uwagę na to że, RNAi jest procesem cytoplazmatycznym, dlatego też siRNA nie może być skierowane na sekwencje intronowe pre-mRNA, którego dojrzewanie zachodzi w jądrze, ale musi być skierowane na kodujące sekwencje genu. siRNA o sekwencjach homologicznych do różnych miejsc mRNA tego samego genu, mogą indukować

zróżnicowany poziom ekspresji genu. Selekcja docelowej sekwencji w siDesigner dokonywana jest za pomocą algorytmów zbudowanych na bazie metod wyszukiwania podobieństw sekwencji. W nowym algorytmie modelowania uwzględniliśmy następujące elementy wpływające pozytywnie na jakość siRNA:

- dobór sekwencji w odległości 50-100 nukleotydów od miejsca inicjacji translacji
- dobór sekwencji siRNA posiadające na końcach 3' reszty urydylowe
- zawartość nukleotydów G-C w projektowanych sekwencjach nie powinna przekraczać 30-40%
- unikanie sekwencji bogatych w G zdolnych do tworzenia struktur czwartorzędowych (ang. tetraplex)
- zastosowaniem mechanizmów analiz tekstowych i zagadnień kluczowych "Go Terms" celu weryfikacji interesującego genu
- eliminacja skutków off-target za pomocą innowacyjnego mechanizmu filtrowania
- klasyfikacja wymodelowanych kandydatów dla docelowego genu z wykorzystaniem metod sztucznej inteligencji

Obserwuje się wyraźną zależność zakresu i poziomu niepożądanych zmian od ilości siRNA użytych do doświadczeń. Istotnym elementem jest uwzględnienie problemu efektów ubocznych poprzez udoskonalanie cząsteczek siRNA (np. na drodze ulepszenia modelowania sekwencji) w taki sposób, aby osiągnąć wysoką skuteczność działania przy minimalnej dawce.

Innym projektem badawczym związanym z modelowaniem RNAi było opracowanie przez nas metod bioinformatycznych klasyfikacji wyników eksperymenalnych oraz zamodelowanie bazy danych esiRNA, na podstawie których byliśmy w stanie zbudować własną bibliotekę RNAi (ang. endoribonuclease-prepared short interfering RNA, esiRNA) (*Kittler R., Kozak K., et al., Nature Methods, 2007*). W następnym etapie, projektu na skalę genomu udało nam się przewidzieć fenotypy, które odpowiadają za podział komórki w procesie mitozy. Zaproponowana przez nas technologia produkcji RNAi bazuje na enzymatycznym trawieniu długiej nici dsRNA za pomocą enzymu RNase III oraz poprzez kombinację procesów interferencji. Udało nam się dotychczas jako jednemu zespołowi utworzyć bibliotekę esiRNA utworzoną na podstawie sekwencji cDNA, które reprezentowały losową część sekwencji transkrypcji. Dzięki bioinformatycznemu systemowi zarządzania i analizy danych, które udało nam się zaprojektować i zaimplementować, stworzyliśmy pakiety narzędzi umożliwiające projektowanie optymalnych cząsteczek RNAi na podstawie biblioteki genomu RNAi (ang. endoribonuclease-prepared short interfering RNAs, esiRNAs) dla człowieka i myszy.

Jednym z głównych celów mojej pracy badawczej jest użycie analiz komputerowych, metod sztucznej inteligencji oraz analiz doświadczalnych do wyciszania genów za pomocą siRNA. W roku 2003 otrzymałem ofertę pracy na stanowisku badawczym w Instytucie Maxa Plancka Biologii Molekularnej i Genetyki (MPI-CBG) w Dreźnie w grupie WSBP. Moim początkowym celem w MPI-CBG było utworzenia systemu i adaptacji metod automatycznego rozpoznawania obrazów w celu analizy mikroskopowych obrazów fluorescencyjnych próbek RNAi i ich oddziaływania na linie komórkowe myszy i człowieka. W pierwszym etapie pracy w zespole WSBP pod kierownictwem Dr Eberharda Krausza dokonaliśmy wielu przełomowych badań i stworzyliśmy pionierskie systemy zarządzania danymi, które umożliwiły nam głębszą analizę funkcjonowania technologii RNAi. Naszym celem było wyszukiwanie docelowych

biologiczno-chemicznych cząsteczek, które odgrywają specyficzną funkcję podczas życia komórki. Pracowaliśmy głównie nad analizą i modelowaniem bibliotek RNAi oraz związkami chemicznymi, skupiając się najczęściej na komórkach człowieka i myszy. Wynikiem pracy badawczej nad metodami klasyfikacji i zarządzania danymi w zespole WSBP, MPI-CBG było ukończenie pracy doktorskiej pt.: „Klasyfikacja danych z biologicznych analiz wieloprzepustowych za pomocą adaptacyjnych metod k-najbliższych sąsiadów”, którą obroniłem w 2007 r.

Podczas kontynuacji badań na uniwersytecie ETH w Zurychu, byłem twórcą statystycznych metod do automatycznej diagnostyki procesów WSBP oraz definicji docelowych genów (publikacja 2 z osiągnięć naukowych, **Kozak K.**: *The Use of Design Specificity in Standardized Mean Difference for Analysis of High throughput RNA Interference Screens. Lecture Notes in Computer Science 7451 Springer 2012, ISBN 978-3-642-32394*). Zaproponowane w tej pracy metody służą przede wszystkim, do automatycznej detekcji docelowych genów. Standardowa średnia różnica (ang. Standardized Mean Difference, SMD) jest metodą służącą do oceny skuteczności wyciszenia siRNA. SMD bazuje na pomiarze błędnie sklasyfikowanych siRNA i eliminacji pseudopozytywnych siRNA. SMD wykorzystuje do tego celu pomiar sygnału próbek kontrolnych (pozytywnych i negatywnych). Pomimo ciągłego ulepszania technologii siRNA, ewolucji transkryptomu, podczas modelowanie cząsteczek siRNA lub podczas procesu syntezy mogą wystąpić cząsteczki, które nie wyciszają żadnego genu (ang. non-targets). Jednym z powodów tego efektu jest zmieniająca się w czasie anotacja genomu i transkryptomu. Powoduje to, że informacja o transkryptach genu podczas wykonywania eksperymentu w porównaniu do informacji z momentu projektowania cząsteczki mogą się różnić. Głównym celem w naszej metodzie SMD było uwzględnienie siRNA „non-target” i potraktowanie tych cząsteczek jako dodatkowe kontrolki negatywne. Istotą projektu było znalezienie bioinformatycznych metod pozwalających na detekcje non-target spośród całej kolekcji zbiorów siRNA. Identyfikacja „non-target” pozwala nam na wprowadzenie nowego parametru w metodzie SMD. Dzięki identyfikacji non-target byłem w stanie zademonstrować wyniki o ilości skutków ubocznych kilku zbiorów siRNA o skali genomu, co pozwoliło nam udostępnić dla laboratoriów informacji o jakości komercyjnie produkowanych siRNA.

Kluczowa część moich badań była poświęcona metodom pozwalającym na pomiar jakości testów biologicznych zorientowanych na WSBP. Udało mi się utworzyć nową metodę badania jakości testu uwzględniającą nieliniowość złożonych danych mikroskopowych (Publikacja 3 z osiągnięć naukowych, **Kozak K., Csucs G.**: *Kernelized Z' Factor in Multiparametric Screening Technology. RNA Biology, 2010, Vol 7, Issue 5, doi: 10.4161/rna.7.5.13239*). Odkrycie zjawiska interferencji RNA zapoczątkowało nowe badania dotyczące hamowania ekspresji wybranych genów przez małe interferujące RNA (siRNA, esiRNA) w komórkach ssaków. Proces wyszukiwania docelowych związków (siRNA, esiRNA) polega na wyszukaniu aktywnych cząsteczek (faworytów) spośród całego genomu. Standardową techniką w odkrywaniu docelowych związków jest ich ewaluacja za pomocą eksperymentu biologicznego. Jak już wstępnie opisałem powyżej, poświęciłem cykl moich publikacji, na skupienie się na analizie i obróbce danych WSBP. Byłem twórcą nowej metody „Kernelized Z Factor” do kontroli testów biologicznych w technologii WSBP. Zaproponowana przeze mnie metoda ilustruje sposób badania jakości testów analiz WSBP biorąc pod uwagę wielowymiarowość zbioru danych. Głównym naszym osiągnięciem było wprowadzenie do metody liniowych analiz

dyskryminacyjnych (LAD) funkcji jądra Gaussa. Wprowadzenie funkcji jądra do metody LAD pozwoliło nam uniknąć problemu w sytuacji gdy poszczególne parametry wyników analiz posiadają rozkład nieliniowy. Dzięki zastosowaniu wprowadzonej metody mogliśmy zbudować test biologiczny dla WSBP polegający na kontroli procesu eksportu kompleksu 60S odpowiadającego za budowę Rybosomu. Dodatkowo, dzięki tej metodzie udało nam się scharakteryzować receptor Exportin 5 (Exp5) odpowiedzialny za export kompleksu 60S w komórkach ludzkich. Stosując techniki siRNA w celu kontroli infekcyjności udało nam się także śledzić transport bakterii Salmonella do wnętrza komórki.

Wynikiem kontynuacji pracy badawczej nad metodami analizy WSBP było opracowanie przeze mnie nowej metody "Z' Factor" uwzględniającej jakość zaprojektowania siRNA (*Mazur S., Kozak K.: Z' Factor including siRNA design quality parameter in RNAi screening experiments, RNA Biology, 2012*). Badania próbek kontrolnych WSBP, są doskonałym uzupełnieniem procesu walidacji i istotnie wspomagają proces ciągłego doskonalenia jakości pomiarów WSBP. Przetworzenie setek próbek i przygotowanie raportów dla próbek kontrolnych zajmuje niejednokrotnie wiele dni i wiąże się z dużym ryzykiem popełnienia błędów. Do oceny jakości eksperymentu na podstawie kluczowych parametrów (np. ilość komórek), które są mierzone w trakcie WSBP stosowane są metody statystyczne. Jedną z najbardziej popularnych metod statystycznych jest współczynnik Z'. Z' wykorzystuje do tego celu pomiar sygnału próbek kontrolnych (pozytywnych i negatywnych). Współczynnik Z' za pomocą porównania rozkładu danych kontrolnych pozytywnych i negatywnych jest w stanie ocenić jakość przeprowadzonego eksperymentu. Im więcej próbek kontrolnych jest częścią analizy, tym eksperyment jest bardziej wiarygodny ale jednocześnie stawia przed naukowcem więcej wyzwań. Zastosowanie większej ilości próbek kontrolnych wiąże się z większymi kosztami projektu WSBP. Podobnie jak w metodzie SMD naszym celem w Z' było uwzględnienie cząsteczek, które nie wyciszają żadnego genu (ang. non-targets). Non-target jest traktowany jako dodatkowa kontrola, przez co daje nam lepsze możliwości oceny jakości eksperymentu bez zwiększania kosztów. Naszym celem było uwzględnienie siRNA non-target i potraktowanie tych cząsteczek jako dodatkowe kontrole negatywne. Istotą projektu było utworzenie metod bioinformatycznych pozwalających na detekcję non-target spośród całej kolekcji zbiorów siRNA. Identyfikacja non-target jest kluczowym elementem procesu i ich precyzyjna detekcja wpływa na cały proces pomiaru jakości WSBP. Do detekcji non-target zastosowaliśmy mechanizm aktualizacji informacji o transkryptach. Identyfikacja non-target pozwoliła nam na wprowadzenie nowego typu kontroli negatywnych w metodzie Z'. Dzięki non-target byliśmy w stanie zwiększyć ilość kontroli oraz zredukować ilość siRNA nie wpływających na poziom wyciszania genów. Dodatkowo dzięki identyfikacji non-target byliśmy w stanie dostarczyć statystyki dostępnych zbiorów siRNA o skali genomu, tym samym udostępnić informacje o jakości komercyjnie produkowanych siRNA.

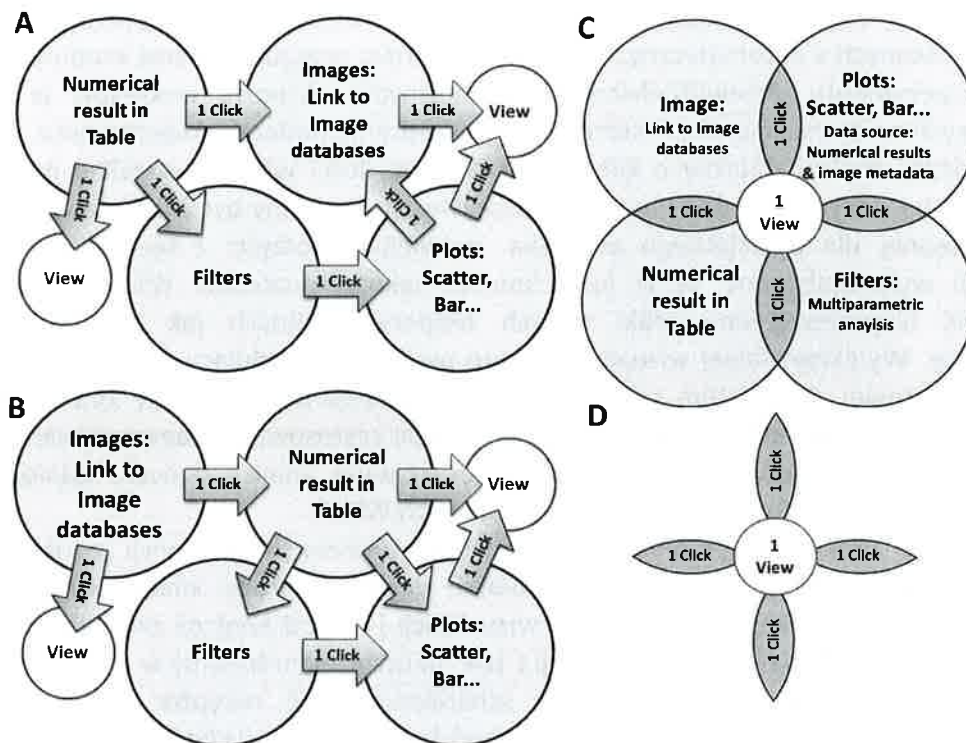
Jak opisałem powyżej do najważniejszych zadań WSBP należy automatyczna ocena jakości eksperymentu za pomocą metod statystycznych. Inną metodą oceny jakości eksperymentu, jest wizualna ocena zbiorów danych z eksperymentu. Wizualizacje ułatwiają zrozumienie skomplikowanych zjawisk biologicznych. W artykule (*Zwolinski L, Kozak M, Kozak K.: 1Click1View: interactive visualization methodology for RNAi cell based microscopic screening. BioMed Research International, Hindawi, 2012*) przedstawiliśmy metodę i zintegrowane środowisko 1Click1View do interaktywnej wizualizacji danych, pochodzących z

eksperymentów WSBP. Zaproponowaliśmy jednolitą i interaktywną reprezentację zbiorów danych pozyskanych z automatycznych mikroskopów oraz przygotowaliśmy moduły do badania jakości eksperymentu. Przewidzieliśmy także możliwość eksportu wykresów w formatach pozwalających na detekcję próbek, które uległy uszkodzeniu podczas eksperymentu.

Podczas analizy zbiorów o kilkunastotysięcznej ilości wierszy, wszelkie mapy, tabele, filtry, wykresy i symulacje próbek oraz dane statystyczne powinny być przedstawione w formie łatwo dostępnej dla przeciętnego analityka. Największe korzyści z technik automatycznej wizualizacji wykorzystywane są w fazie kontroli jakości procesów WSBP. Pozwalają one przedstawić nieprzetworzone bloki danych eksperymentalnych jak i surowe rezultaty obliczeniowe. Wynikiem takiej wizualizacji może być obraz znajdujący się w przestrzeni dwu- lub trójwymiarowej. Rezultatem naszej pracy było udostępnienie metody 1C1V używanej do analizy i wizualizacji obrazów fluorescencyjnych. Dzięki zastosowaniu zaawansowanych technik obrazowania i wyrafinowanych funkcji do interaktywnej analizy danych, udało nam się wyszukiwać klasy podzbiorów z ogromnych baz danych WSBP.

Podczas pracy eksperymentalnej, studiując procesy interferencji z użyciem RNAi, skupialiśmy się w szczególności na podzbiorze genów z grupy kinaz. Stosując technikę wyciszania RNAi i zaawansowane techniki wizualizacji i metod kontroli próbek eksperymentu wraz z metodami automatycznej klasyfikacji z tzw. nauczycielem byliśmy w stanie monitorować formowanie rybosomu. Udało nam się scharakteryzować receptor Exportin 5 (Exp5) odpowiedzialny za export kompleksu 60S w komórkach ludzkich. (*Wild T., Horvath P., Wyler E., Widmann B., Badertscher L., Zemp I., Kozak K., Csucs G., Lund E., Kutay U.: A Protein Inventory of Human Ribosome Biogenesis Reveals an Essential Function of Exportin 5 in 60S Subunit Export. PLoS Biol, 2010*). Zademonstrowaliśmy, że Exp5, w podobieństwie do dotychczas znanej 60S exportin Crm1, łączy się z cząsteczkami pre-60S za pośrednictwem RanGTP. Zastosowanie technologii RNAi i wyciszanie białek kompleksu zawierającego Exp5 lub Crm1 powoduje blokowanie exportu 60S zarówno w komórkach ludzkich i w żabich, a tym samym ma wpływ na formowanie rybosomu. Natomiast export 40S jest zablokowany poprzez CM1. Dzięki praktycznemu zastosowaniu moich metod, podczas eksperymentu kontroli exportu kompleksu 60s odpowiedzialnego za produkcję rybosomu, udało mi się zidentyfikować siRNA dla genu DTNBP1, którego wyciszenie dało nieoczekiwany fenotyp.

W celu monitorowania zachowania komórek w procesie biogenezy, zastosowaliśmy 1C1V do opracowania kilku niestandardowych aplikacji, które zarządzają ogromnymi ilościami danych z mikroskopów konfokalnych i umożliwiają tworzenie filmów śledzących ruchy pojedynczych komórek. Zdecydowaliśmy się na rozwój 1C1V, ponieważ szukaliśmy narzędzia, które dawałoby dostęp do niezbędnych funkcji matematycznych, a także pozwalałoby realizować codzienne zadania, takie jak wyświetlanie obrazu bez pisania dużej ilości kodów. Pisane pod użytkownika aplikacje laboratoryjne umożliwiają pracującym biologom wyizolowanie, obcięcie i obróbkę danych, a także wybranie podzbiorów tych danych, rekonstrukcje danych do modeli dwuwymiarowych, a ostatecznie wygenerowanie obrazów śledzących rozwój i zachowanie się komórek. Nasz zespół dalej rozwija 1C1V i poszerza listę zastosowań w aplikacjach WSBP. 1C1V jest narzędziem, które pozwala wyświetlić dane różnego typu na jednej mapie (Rysunek 2).



Rysunek 2: A i B – tradycyjny analityczny scenariusz dla zbiorów danych zawierających obrazy. Wielokrotne kliknięcia powodują generowanie wielu masek. C. Metoda 1C1V: zastosowanie jednej maski i dostęp do danych, narzędzi analitycznych (obrazy, wykresy, wzniki analiz, interaktywne filtry). D. Model gwiazdy jako metodologia oprogramowania do interaktywnej wizualizacji.

Powstały firmy biotechnologiczne specjalizujące się w projektowaniu i wytwarzaniu siRNA i esiRNA, wykorzystywanych do badań w obszarze genomiki funkcjonalnej, a więc do odnajdywania nowych funkcji genów, poznawania mechanizmów sygnalizacji komórkowej, apoptozy, proliferacji czy różnicowania się komórek. Zauważono także olbrzymi potencjał cząsteczek siRNA do „wyłączania” genów związanych z patogenezą chorób. Niektóre z tych firm dostarczają biblioteki cząsteczek RNAi do wyciszania dowolnego genu: Ambion, Dharmacon, Qiagen. Cząsteczki siRNA stały się więc uniwersalnymi, sekwencyjnie specyficznymi narzędziami, pozwalającymi z jednej strony na identyfikację celów terapeutycznych, a więc genów, których wyciszenie prowadzi do cofnięcia lub zatrzymania choroby, a z drugiej – na weryfikację tego wyboru jako związków o potencjalnym zastosowaniu terapeutycznym.

Pomimo ciągłego ulepszania technologii siRNA, oraz pomimo 10-cio letniego okresu rozwoju algorytmów służących do projektowania sekwencji tych cząsteczek, nadal, podczas procesu syntezy występują błędy. W mojej pracy badawczej (*Kozak K.: Annotation and specificity of existing genome-wide siRNA libraries. Nucleic Acid Therapeutics, 10.1089/nat.2012.0387, 12*) udowodniliśmy, że jednym z powodów błędnego modelowania cząsteczek jest wpływ ewolucji baz danych i anotacji transkryptomu. Transkryptom (RefSeq) jest

podstawowym źródłem i niezbędną bazą, używaną w modelowaniu siRNA. Różnica informacji o transkryptach genu podczas wykonywania eksperymentu w porównaniu do informacji z momentu modelowania cząsteczki może mieć istotny wpływ na jakość wyciszenia i identyfikację docelowego genu. W mojej pracy udowodniłem, że ewolucja anotacji genomu wpływa także na proces walidacji docelowych związków. W pracy badawczej nad specyfikacją bibliotek genomu, wskazaliśmy, które parametry algorytmów modelowania siRNA wpływają pozytywnie na proces wyciszania, a które negatywnie. Charakterystyka wieloparametryczności algorytmów modelowania siRNA ma istotny wpływ na jakość projektowanych siRNA w zależności od używanej biblioteki. Cząsteczki wygaszające ten sam gen a pochodzące od różnych firm lub wymodelowane innym algorytmem, posiadają różną jakość wyciszenia. Dzięki metodom porównawczym byliśmy w stanie wskazać parametry modelowania, które są parametrami stałymi, czyli stosowanymi przez większość istniejących algorytmów modelowania. Efektem naszych analiz było zademonstrowanie zbioru parametrów, które powinny być podstawą każdego algorytmu modelowania. Dodatkowo, wskazaliśmy, że identyfikator genu lub jego symbol nie jest prawidłowym kluczem komunikacyjnym w pracach eksperymentalnych i opisowych. Efekty naszej pracy wprowadziły w technologii WSBP nowy element procesu, który umożliwił dokładniejszą weryfikację sukcesu projektów.

Bardzo istotnym etapem badań dotyczącym problemu anotacji bibliotek siRNA i esiRNA było utworzenie bioinformatycznego systemu baz danych pozwalającego na śledzenie ewolucji transkryptomu. Głównym osiągnięciem projektu (*Mazur S., Csucs G., Kozak K.: RNAiAtlas: a database for RNAi (siRNA) libraries and their specificity, Oxford Bioinformatics, 2012*) było zbudowanie ogólnodostępnej bazy danych RNAiAtlas dla cząsteczek RNAi i ich anotacji. Szczególnym osiągnięciem utworzonej pierwszej w świecie ogólnodostępnej oraz darmowej bazy danych dla bibliotek RNAi, było umożliwienie kontrolowania jakości poszczególnych produktów RNAi dostarczanych przez przedsiębiorstwa komercyjne. Wyszukiwarka, wizualizacja i łatwy dostęp do wszystkich danych umożliwił indywidualną i wieloskalową walidację docelowych siRNA. Model i metody zastosowane w RNAiAtlas umożliwiają aktualizowanie anotacji poszczególnych cząsteczek. Dzięki modelowi bazy danych, który pozwala na uzyskanie na jednej karcie wszystkich istniejących cząsteczek na rynku dla danego genu, uzyskaliśmy możliwości ewaluacji i doboru cząsteczek o najwyższej jakości wyciszenia. Karta danych w RNAiAtlas przedstawia cząsteczki siRNA oraz ich parametry funkcjonalności, parametry wyników analiz bioinformatycznych takie jak: ilość docelowych transkryptów, on/off-target, docelowy GeneID oraz właściwości algorytmu modelowania. Jednym z najważniejszych osiągnięć w projekcie RNAiAtlas był nie tylko dostęp do parametrów cząsteczek ale także możliwość monitorowania tych parametrów w czasie. Biolodzy są w stanie za pomocą prostej nawigacji sprawdzić jakie są relacje pomiędzy różnymi wersjami anotacji transkryptomu. Zademonstrowaliśmy, że podstawowy parametr jakim jest Symbol Genu, jego synonimy i połączenie z identyfikatorem GeneID, ulega zmianie w czasie. Z naszych analiz wynika, że Symbol Genu w okresie 5 lat ewolucji anotacji genomu uległ zmianie dla 30% genów. Innym istotnym parametrem, który drastycznie wpływa na modelowanie siRNA, jest niestabilność ilości transkryptów dla danego genu. Informacja o ilości transkryptomów, na bazie których zostało zaprojektowane siRNA, względem istniejących transkryptomów w bazie danych dla danej wersji genomu jest kluczowa. Dodatkowo RNAiAtlas zawiera, krótkie sekwencje wszystkich siRNA i

pozwała na eksport całych bibliotek i ich anotacji. Baza danych jest zespólna z systemem OpenBis, utworzonym dla monitorowania procesów WSBP, która dostępna jako platforma „Open source” pod adresem internetowym: <http://www.cisd.ethz.ch/software/openBIS> (**Kozak K., et al. European Pharmaceutical Review, 2010**).

Poza mikroskopowymi projektami WSBP dotyczącymi wyciszenia ekspresji genów byłem także twórcą narzędzia HCDC do automatycznej analizy obrazów fluorescencyjnych http://hcdc.ethz.ch/index.php?option=com_content&view=article&id=11&Itemid=10. Narzędzie HCDC służy przede wszystkim, do automatycznej analizy fluorescencyjnych obrazów pochodzących z opisanych powyżej projektów. Celem analizy jest segmentacja i detekcja komórek oraz innych obiektów uzyskanych za pomocą mikroskopów automatycznych. Szczególnie owocna okazała się możliwość korzystania z dostępnych algorytmów automatycznej segmentacji obrazów fluorescencyjnych w formie „workflow”. Innym kluczowym zagadnieniem zastosowania analizy danych z zastosowaniem bioinformatycznego środowiska workflow, jest połączenie automatycznej analizy obrazów z wieloklasową klasyfikacją komórek. Celem mojego projektu podczas współpracy z grupą Prof. Wolf-Dietrich Hardta był rozwój metod do automatycznej wieloklasowej klasyfikacji w celu identyfikacji złożonych fenotypów. Efektem naszej pracy było zbudowanie narzędzia Enhance Cell Classifier (ECC), dostępnego na stronie internetowej: <http://yacc.ethz.ch> (*Misselwitz B., Kozak K. Et al. BMC Bioinformatics 2010*). Narzędzie ECC służy przede wszystkim, do automatycznego uczenia zbiorów danych pochodzących z obrazów oraz przeprowadzenia klasyfikacji za pomocą metod bazujących na funkcji jądra. Dzięki ECC utworzyliśmy, jako jedni z pierwszych, mikroskopowy test biologiczny dla pofalowanych komórek stosując do eksperymentów komórkę HeLa. Udało nam się dotychczas jako jednemu zespołowi utworzyć za pośrednictwem ECC analizę procesu dokowania Salmonelli do komórek mitotycznych. Celem projektu była obserwacja inwazji bakterii Salmonella do komórek ludzkich. Moim wkładem w ten projekt była praca nad funkcjami jądra klasyfikatora i samymi klasyfikatorami.

Za najważniejsze osiągnięcia naukowe uważam:

1. Opracowanie zaawansowanych metod modelowania cząsteczek siRNA i kontroli ich skutków ubocznych
2. Opracowanie skutecznych metod oceny jakości WSBP z zastosowaniem RNAi.
3. Opracowanie skutecznych metod wyszukiwania docelowych związków RNAi z eksperymentów WSBP.
4. Opracowanie atlasu RNAi zawierającego aktualne anotacje i metadane RNAi

5. Omówienie pozostałych osiągnięć naukowo-badawczych.

5.1 Oryginalne opublikowane prace twórcze po doktoracie (wkład własny opisany w wykazie):

1. **Kozak K.:** Herausforderung durch hohe Datenmengen aus OLED-Biosensoren. *Meditronic-Journal*, 4/2013

2. Shaoli D., Ghosal S., Chakrabarti J. and **Kozak K.**: SeedSeq - off-target transcriptome database. *BioMed Research International*, 2013.
3. Biernatowska A., Podkalicka J., Majkowski M., Hryniewicz-Jankowska A., Augoff K., **Kozak K.**, Korzeniewski J, Sikorski A.F.: The role of MPP1/p55 and its palmitoylation in resting state raft organization in HEL cells. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2013, March, DOI: 10.1016/j.bbamcr.2013.03.009
4. Michel P., Aloyse A., Trefois C., Stojanovic A., Sagatovich Baumuratov A., **Kozak K.**: Light microscopy applications in systems biology: opportunities and challenges. *Cell Communication and Signaling*, 2013, 11:24 doi:10.1186/1478-811X-11-24
5. **Kozak K.**, Stapor K.: SiRNA sequence model: redesign algorithm based on available genome wide libraries. *Journal of Biomolecular Structure and Dynamics, J Biomol Struct Dyn.* 2012 Dec 20.
6. Hamecher J., Riess T., Bertini E., **Kozak K.**, Kastl J., Mayer T.U, Merhof D.: A versatile framework for the analysis of high-throughput screening data, *Computational Biology*, 2011
7. **Kozak K.**, Csucs, G.: HCIP- imageJ in workflow based environment. *Computational Biology*, 2011
8. **Kozak K.**, Kaestner S., Wild T, Vonderheit A, Misselwitz B, Kutay U, Csucs G: New Generation Algorithm For Analysis Of Off-target Effects In Sirna Screens Paper. Springer-Verlag in a CCIS Series book, 2013
9. Davies, A., **Kozak K.**, Djaballah, H., Arkin, M., Lemmon, V., Turpin, P., Zock, J.: High Content Analysis Roundtable, *International Drug Discovery*, 2011
10. Wild T., Horvath P., Wyler E., Widmann B., Badertscher L, Zemp I., **Kozak K.**, Csucs G., Lund E, Kutay U.: A protein inventory of human ribosome biogenesis reveals an essential function of Exportin 5 in 60S subunit export, *Plos Biology*, 2010
11. **Kozak K.**, Bakos G., Hoff A., Bennett E., Dunican D., Davies A., Kelleher D., Long A., Csucs G.: Workflow based software environment for large-scale biological experiments. *Journal of Biomolecular Screening, Special Issue – HCS*, doi:10.1177/1087057110377354
12. **Kozak K.**, Bauch B., Csucs G., Pylak T. & Rinn B.: Towards a comprehensive open source platform for management and analysis of High Content Screening data, *European Pharmaceutical Review*, Issue 4, 08.2010
13. **Kozak K.**, Agraval, A., Machuy, N., and Csucs, G. (07/2009): Data Mining Techniques in High Content Screening. *J Comput Sci Syst Biol* 2: 219-239. doi:10.4172/jcsb.1000035.
14. Misselwitz. B , Strittmatter B., Periaswamy,B., Schlumberger, M. C, Rout, S., Horvath, P., **Kozak K.**, Hardt W., D.: CellClassifier: a multi-class classification tool for microscopy images . *BMC Bioinformatics*, 2010;11:30
15. **Kozak K.**, Firkowski A., Csucs, G.: Open source data management in High Content Screening technology. *European Pharmaceutical Review*, 2009, 02, 16-24
16. Kittler R., Pelletier L, Heninger A.K., Slabicki M., Theis M., Miroslaw L, Poser I, Lawo S., Grabner H., **Kozak K.**, Wagner J., Surendranath V., Richter C., Bowen W., Jackson A.L., Habermann B., Hyman A.A., Buchholz F.: Genome-scale RNAi profiling of cell division in human tissue culture cells, *Nature Cell Biology*, 2007, 9, 1401 – 1412.

5.2 Oryginalne opublikowane prace twórcze przed doktoratem:

17. Heninger A.K, Kozak K., Kittler R., Wagner J., Lohmann A., Poser I., Grabner H., Krausz E., Buchholz F.: RNAi Libraries for Functional Genomics. Avoid the Difficulties of Predicting Efficient siRNAs and the Cost of Chemical Synthesis. GEN Nov, 2006, Vol. 26, No. 19, 12-16
18. Kittler, R., Surendranath, V., Heninger, A.K., Slabicki, M., Theis, M., Putz, G., Franke, K., Caldarelli, A., Grabner, H., **Kozak, K.**, Wagner, J., Rees, J., Korn, B., Sachse, C., Soennichsen, B., Guo, J., Schelter, J., Burchard, J., Linsley, P., Jackson, A.L., Habermann, B. and Buchholz, F. (2007): Genome-wide Resource of Endoribonuclease-prepared short interfering RNA for specific loss-of-function studies in Human, Mouse and Rat. Nature Methods, 4, 337-44: A
19. **Kozak K.**, Eshun B. and Oegema J.: The handling and analysis of large scale high content screening data. European Pharmaceutical Review, 2007, 3, 31-35.
20. **Kozak K.**: Automated Data Flow and Quality Control Practice in High-Content Screening, Saechsische Biotechnologiebuch, Dresden Verlag, 2007, 11
21. **Kozak K.**, Kozak M., Stapor K.: Optimization of the SVM Screening Kernel. Application to Hit Definition in Compound Screening. Computer Recognition Systems, Springer, 2007, 224-231 , DOI10.1007/978-3-540-75175-5 2
22. **Kozak K.**, Kozak M., Stapor K.: Kernels for chemical compounds in biological screening, Springer Verlag, Computer Science, 2007, DOI10.1007/978-3-540-71629-7_37, 327-337
23. **Kozak K.**, Kozak M., Stapor, K.: Weighted k-Nearest-Neighbor Techniques for High Throughput Screening Data. International Journal of Biomedical Sciences, 2006, 155 - 160.
24. Grabner H., Heninger A. K, Kittler R., Lohmann A., Poser I., Wagner J., **Kozak K.**, Buchholz F., Krauß E.: Production of a genome-wide RNA interference library for functional gene characterization in cultured human cells. Tecan Journal 2006, vol 2 ,10-12
25. **Kozak K.**, Kozak M., Krausz, E.: SIB: database and tool for the integration and browsing of large scale image high-throughput screening data. IEEE, DEXA, 2006, pp.201-205
26. Proszynski T.J., Klemm R., Gravert M., Hsu P., Wagner J., **Kozak K.**, Grabner H., Habermann H., Bagnat M., Simons K., Walch-Solimena K.: A visual screen for sorting mutants in yeast biosynthetic pathways using the systematic deletion mutant array. PNAS, 2005, December 13; 102(50): 17981–17986
27. Krausz, E., Heninger, A.K., Kittler, R., Lohmann, A., Poser, I., Wagner, J., **Kozak, K.**, Grabner, H. And Buchholz, F.: Eine Genom-weite RNA-Interferenz-Bibliothek für die funktionelle Charakterisierung aller Gene in kultivierten, menschlichen Zellen (in german). BIOSpektrum 2005; 11:436-441.
28. Heninger, A.K., Franke, K., Kittler, R., **Kozak, K.**, Wagner, J., Slabicki, M., Ding, L. and Buchholz, F.: Generation of mouse esiRNA library, FunGenEs, 2005, 44-51
29. **Kozak, K.**: Methods for increasing of information search quality – automatic categorization. Study of Informatics, 2004, Volume 25, No. 3, 59-69



5.3 Streszczenie wyników uzyskanych w ww. publikacjach

W latach 2000-2001 pracowałem w firmie Deutsche Bahn gdzie zajmowałem się programowaniem baz danych. Następnie w roku 2001 zostałem przyjęty do głównej siedziby Instytutu Maxa Plancka (MPI) w Monachium. Tam zacząłem badania w ramach pracy doktorskiej, której przedmiotem było zastosowanie metody rozpoznawania obrazów, klasyfikację tekstu i metody uzyskiwania informacji dla złożonych zbiorów dokumentów tekstowych. Zajmowałem się przede wszystkim projektowaniem modelu meta-danych „dublin core” jak również budową systemu pozwalającego na obróbkę tych danych. W tym czasie centrala MPI zwróciła się do naszej grupy z prośbą o zbudowanie systemu zarządzania publikacjami naukowymi poszczególnych jednostek naukowych Stowarzyszenia Maxa Plancka. Efektem naszej pracy było utworzenie systemu eDoc www.edoc.mpg.de. Głównym celem systemu eDoc jest zarządzanie danymi typu tekstowego. Jądro tego systemu stanowią manuskrypty, dane numeryczne oraz dane graficzne opisujące wyniki biologiczne z biomedycznych MPI.

W trakcie projektu w Monachium prezentowałem swoje wyniki na sympozjum w Lipsku (Niemcy), a także zostałem autorem publikacji w Zeszytach Politechniki Śląskiej (*Kozak K. Study of informatics, 2004*). Nasze główne zadanie polegało na analizie informacji z publikacji naukowych oraz z zawartości obrazów cyfrowych, w celu wyszukiwania udokumentowanych funkcji genów. Zadanie odkrywania funkcji genów oraz ich wpływ na życie komórek różnych organizmów wydało mi się wyjątkowo fascynujące. W szczególności interesował mnie rozwój analiz komputerowych i metod sztucznej inteligencji do analizy danych doświadczalnych tej dziedziny.

W roku 2003 otrzymałem ofertę pracy na stanowisku badawczym w Instytucie Maxa Plancka Biologii Molekularnej i Genetyki (MPI-CBG) w Dreźnie w grupie WSBP. Moim początkowym celem w MPI-CBG było utworzenia systemu i adaptacji metod automatycznego rozpoznawania obrazów w celu analizy mikroskopowych obrazów fluorescencyjnych komórek, w których stosowano wyciszanie poszczególnych genów za pomocą RNAi.

Wynikiem pracy badawczej w zespole WSBP nad metodami klasyfikacji danych oraz metodami ich zarządzania było ukończenie przeze mnie pracy doktorskiej pt.: „Klasyfikacja danych z biologicznych analiz wieloprzepustowych za pomocą adaptacyjnych metod k-najbliższych sąsiadów” w celu ukończenia pracy doktorskiej w dziedzinie informatyki (pracując nad badaniami w instytucie biologii), skierowałem prośbę do Wydziału Informatyki Politechniki Śląskiej o ewaluację mojego dorobku. Po wyrażeniu zgody przez władze wydziału Informatyki Politechniki Śląskiej obroniłem pracę doktorską pod opieką Prof. dr hab. inż. Katarzyny Stąpor, która jest pionierem i ekspertem w dziedzinie sztucznej inteligencji i automatycznego rozpoznawania obrazów. Zasadniczym celem pracy doktorskiej było znalezienie relacji pomiędzy strukturą związku chemicznego i jego biologiczną aktywnością w komórkach rakowych, z założeniem, że podobieństwo struktur cząsteczek chemicznych wykazuje podobieństwo w biologicznych procesach. W celu badania aktywności antyrakowych przeanalizowałem struktury 3000 związków chemicznych biorących udział w dotychczas przeprowadzonych eksperymentach

WSBP. Głównym celem WSBP była ewaluacja syntetycznych cząsteczek lub naturalnych próbek ukazując selektywne zahamowanie podziału komórki w poszczególnych liniach nowotworowych (białaczka, czerniak złośliwy, rak płuc, okrężnica, mózg, prostata, rak nerek). Głównym osiągnięciem mojej pracy doktorskiej, było zamodelowanie nowej funkcji jądra dla klasyfikatora wykorzystywanego w technologii WSBP, która ma bardzo duży wpływ na precyzyjność klasyfikacji danych. Jądro spełnia szereg ważnych funkcji, z których najważniejsze stanowią bazę dla podstawowych elementów klasyfikatorów typu „maszyna wektorów nośnych” (ang. support vector machine, SVM) czy adaptacyjnych k - najbliższych sąsiadów (ang. k- nearest neighbors, knn) powszechnie stosowanych do wieloparametrycznych analiz bioinformatycznych.

W roku 2006 zostałem poproszony przez dyrektora instytutu MPI-CBG o pokierowanie grupą bioinformatyczną zajmującą się budową systemów zarządzania danymi oraz analizą danych z WSBP, z zastosowaniem technologii RNAi. Jednym z efektów pracy naszego zespołu było utworzenie systemu zarządzania danymi laboratoryjnymi, który znalazł zastosowanie w wielu instytucjach zajmujących się badaniami genetycznymi, między innymi w Międzynarodowym Instytucie Biologii Molekularnej i Komórkowej w Warszawie (IIMCB).

Alternatywną metodą wyciszania genów za pomocą RNAi jest generowanie poprzez enzymatyczną hydrolizę *in vitro* komplementarnych do fragmentu docelowego genu całej puli długich cząsteczek dwuniciowego RNA (Kittler i wsp. 2004) czyli tzw. esiRNA, które są skierowanych na jeden docelowy gen. Jakkolwiek takie podejście gwarantuje największą skuteczność wywołania pożądanego efektu, „wąskim gardłem” jest zawsze precyzyjny dobór specyficznej sekwencji docelowej. Głównym celem mojej pracy badawczej w MPI-CBG było utworzenie wspólnie z zespołem Dr Franka Buchholz biblioteki esiRNA na skali całego genom człowieka (Kittler R., Kozak K., et al. *Nature Methods*, 2007). Tworzenie biblioteki przebiegało etapami: w pierwszej fazie udało nam się uzyskać esiRNA dla kinaz a w następnej fazie dla całego genomu człowieka. Po uzyskaniu biblioteki esiRNA dla genomu człowieka, stosując te same techniki, przeprowadziliśmy produkcję biblioteki esiRNA dla genomu myszy. Projekt zakończył się przeprowadzeniem eksperymentu na skalę genomu w celu wyselekcjonowania esiRNA, wpływającego na proces podziału komórki (Kittler R., Kozak K., et al. *Nature Cell Biology*, November 2007). Na wszystkich etapach tych badań mój udział polegał głównie na analizie i kontroli przepływu danych pomiędzy mikroskopem a analizatorem obrazów, udostępnieniu narzędzi i algorytmów automatycznej obróbki obrazów cyfrowych, analizie bioinformatycznej oraz funkcjonalnemu grupowaniu otrzymanych fenotypów. Zarówno moje własne badania, które wykonałem w ramach pracy doktorskiej jak i późniejsza praca w MPI-CBG pozwoliły mi pogłębić wiedzę w zakresie genetyki, budowy cząsteczek RNAi oraz mechanizmu wyciszania RNAi. Wszystkie te elementy przekonały mnie o szerokich możliwościach technologii RNAi w poznawaniu nowych funkcji genów. Dzięki opublikowaniu pracy dotyczących pionierskich platform bioinformatycznych dla WSBP udało mi się nawiązać wiele kontaktów ze środowiskiem pracującym nad technologiami RNAi oraz nawiązać współpracę z jednostkami zajmującymi się badaniami automatycznego procesu wyciszania ekspresji genów.

Wyniki swoich badań w zakresie analizy danych RNAi, wielokrotnie prezentowałem na konferencjach o zasięgu międzynarodowym. W roku 2005 zostałem zaproszony do przyłączenia się do organizacji Cambridge Healthcare (znajdującej się w Stanach Zjednoczonych), która zajmuje się rozpowszechnianiem standardów i technologii związanej z WSBP. Swoją działalność

dydaktyczną i doradczą prowadzę do dziś a zakres materiałów oraz poziom prowadzonych kursów poszerzam na nabieżąco.

Dr Gabor Csucs, dyrektor centrum mikroskopowego (LMC) na Technicznym Uniwersytecie w Zurychu (ETH), był jednym z uczestników konferencji WSBP organizowanej przez moją grupę w MPI-CBG w roku 2006. Po spotkaniu z Dr Csucs, otrzymałem propozycję pracy na stanowisku kierownika bioinformatycznej grupy zarządzania danymi przy nowo tworzonej jednostce WSBP na ETH. Początkowo pracowałem równoległe dla MPI-CBG i ETH, później podjąłem pracę na pełnym etacie na ETH, w nowej jednostce naukowej, gdzie kontynuowałem pracę nad metodami automatycznej klasyfikacji RNAi w celu zastosowania w technologii WSBP. Wiosną 2008 roku podczas mojej pracy badawczej na ETH założyłem projekt, w którym postawiałem sobie ambitne zadanie: stworzenie w pełni ogólnodostępnej i darmowej platformy, która umożliwiałaby zarządzanie bibliotekami RNAi, uporządkowanie danych, ich wizualizację, segmentację obrazów mikroskopowych oraz zawierała elementy bioinformatyki. Platforma taka powstała w roku 2009 i była między innymi wykorzystana do eksperymentu WSBP, którego celem była identyfikacja nowych białek zaangażowanych w mechanizm migracji komórki T. W ramach tego projektu mierzyliśmy zmiany morfologiczne komórki T z linii komórkowej chłoniaków (Hut78) podczas aktywacji receptora LFA1 stosując przeciwciało SPVL-7. Analiza polegała na obserwacji 1109 genów, a każda próbka posiadała 7500 komórek. Głównym celem postawionym temu projektowi był rozwój modułów, które posiadają łatwy dostęp do narzędzi automatycznej klasyfikacji danych. Efektem naszej pracy było utworzenie platformy High Content Data Chain (HCDC) (Kozak K., Bakos G., et al.: *Workflow based software environment for large-scale biological experiments. Journal of Biomolecular Screening, Special Issue – HCS, 2009, Kozak K. and Csucs, G. European Pharmaceutical Review, 2009*).

Od momentu udostępnienia platformy jako darmowego narzędzia, grono użytkowników stale się powiększa a przyłączenie się innych programistów sprawia, że platforma HCDC ciągle się rozwija. Platforma jest dostępna pod adresem: <http://hcdc.ethz.ch>. Budując HCDC udało nam się utworzyć metodologię podejścia do analizy danych RNAi polegającą na wizualnym projektowaniu procesu analizy. Obecnie nasza platforma jest zainstalowana w co najmniej 20 jednostkach naukowych i firmach farmaceutycznych. Platforma HCDC dała nam możliwość integracji wyników projektu badawczego dotyczącego analizy efektów ubocznych wyciszenia ekspresji genów za pomocą siRNA. Podczas badań nad efektami ubocznymi, korzystając z algorytmów bioinformatycznych udało nam się udowodnić, że siRNA poprzez podobieństwo sekwencji wykazują efekt docelowy na inną sekwencję mRNA niż oryginalnie wymodelowane mRNA (efekt uboczny wyciszenia). Stworzyliśmy szereg algorytmów na bazie narzędzi BLAST, dopasowania sekwencji, filtrów analizy tekstowej oraz pozycji sekwencji zasiania siRNA w celu inspekcji efektów ubocznych. Jako przykład, udało nam się zidentyfikować siRNA dla genu DTNBP1, które nieoczekiwanie wyciszyło ekspresję innego genu. Jedno z czterech siRNA (1/4) genu DTNBP1 wyciszało nieoczekiwanie jako efekt uboczny inną sekwencję mRNA niż tą oryginalnie zamodelowaną (RPS3). Podobny efekt uzyskaliśmy dla siRNA genu OVOS2. Za pomocą algorytmów podobieństwa sekwencji oraz analizy tekstowej, udało nam się udowodnić, że siRNA genu DTNBP1 wycisza mRNA genu RPS3 (fenotyp wyciszenia DTNBP1 był potencjalnie najbliższy fenotypowi otrzymanemu w eksperymencie wyciszenia genu RPS3). Za pomocą algorytmów badania podobieństwa sekwencji jesteśmy w stanie przewidzieć, że nie istnieje żadne podobieństwo z innym mRNA i możemy traktować sygnał takiego siRNA w

eksperymentach jako efekt uboczny. W eksperymencie dotyczącym inwazji bakterii Salmonella do komórek, także przeprowadziłem analizę skutków ubocznych siRNA 1/4 lub 2/4. Dzięki analizie mogliśmy przewidzieć, że siRNA genu KCNC1, generuje efekt uboczny co ma bardzo duży wpływ na blokowanie inwazji bakterii.

Dodatkowo dużą część moich badań poświęciłem metodom pozwalającym na pomiar jakości testów biologicznych zorientowanych na WSBP. Udało mi się utworzyć nową metodę badania jakości testu uwzględniającą nieliniowość złożonych danych mikroskopowych. Poza mikroskopowymi projektami WSBP dotyczącymi wyciszania ekspresji genów byłem także twórcą narzędzi do automatycznej analizy obrazów fluorescencyjnych http://hcdc.ethz.ch/index.php?option=com_content&view=article&id=11&Itemid=10 (Kozak, K., Csucs, G.: HCIP- imageJ in workflow based environment. *Computational Biology*, 2011). Narzędzia te służą przede wszystkim, do automatycznej analizy obrazów fluorescencyjnych pochodzących z opisanych powyżej projektów. Celem analizy jest segmentacja i detekcja komórek oraz innych obiektów uzyskanych za pomocą mikroskopów. Szczególnie owocna okazała się możliwość korzystania z dostępnych algorytmów automatycznej segmentacji obrazów fluorescencyjnych w formie workflow.

Podczas mojej pracy na ETH zostałem zaproszony do poprowadzenia projektu i kierowania zespołem bioinformatyków i inżynierów oprogramowania. Grupa została utworzona dzięki finansowaniu, które pochodziło ze szwajcarskiej fundacji SystemX. Uzyskane finansowanie wykorzystujemy do celów badań w technologii WSBP. Głównym celem projektu było zbudowanie ogólnodostępnej bazy danych do zarządzania danymi typu obraz fluorescencyjny oraz do analizy danych numerycznych. Szczególnym osiągnięciem utworzonej pierwszej w świecie darmowo dostępnej bazy danych dla WSBP, było umożliwienie szybkiej analizy danych obrazowych, gromadzenie i wizualizacja efektu wyciszania genów w komórce. Wizualizacja i łatwy dostęp do obrazów umożliwił manualną walidację docelowych siRNA. Baza danych jest zespólna z wyżej wymienionym narzędziem HCDC i dostępna jako platforma „Open source” pod adresem internetowym: <http://www.cisd.ethz.ch/software/openBIS> (Kozak K., et al. *European Pharmaceutical Review*, 2010).

Planując eksperyment WSBP należy oprócz ustalenia platformy i rodzaju WAG odgórnie dokonać wyboru metod analizy, dostosowując je do problemu biologicznego, który chcemy rozwiązać. Biorąc pod uwagę trudności dotyczące doboru metod analizy danych postanowiłem, wspólnie z innymi współautorami pracy napisaliśmy pracę przeglądową, opisującą działanie wielu algorytmów i etapów analizy (Kozak K., et al. *J Comput Sci Syst Biol*, 2009). W naszym projekcie przeglądowym skupiliśmy się na wykorzystywaniu algorytmów uczących się do celów obserwacji danych RNAi.

W kolejnym projekcie badawczym na ETH skupiłem się na strukturze danych związków chemicznych mających wpływ na aktywność komórki w WSBP (Kozak K., et al. *Springer Verlag, Computer Science*, 2007). Pracowaliśmy nad rozwojem specyficznej dla tych danych funkcji jądra, która odgrywa bardzo istotną rolę w klasyfikatorach zbiorów danych chemicznych. Kluczową rolę w klasyfikacji danych chemicznych odgrywały wygenerowane deskryptory cząsteczek chemicznych. Ostatecznym wynikiem mojej pracy badawczej było utworzenie funkcji jądra dla związków chemicznych na bazie ich dekryptorów, która podobnie jak jądra graficzne, pomaga nie tylko usprawnić klasyfikację danych WSBP. Dodatkowo dzięki implementacji deskryptorów optymalizuje proces analizy dla tego typu danych.

Oprócz działalności naukowej zajmuję się także dydaktyką i popularyzacją bioinformatyki wśród „biologów-doświadczalników”. Prowadzę regularne wykłady na Uniwersytecie Drezdeńskim oraz jestem zapraszany do organizowania 1 lub 2-dniowych kursów/warsztatów analizy danych obrazów mikroskopowych, adresowanych głównie dla starszych studentów oraz młodych jak i doświadczonych pracowników naukowych. Takie kursy, składają się zwykle z kilku godzin wykładów i/lub kilku godzin ćwiczeń.

5.4. Kierowanie międzynarodowymi lub krajowymi projektami badawczymi lub udział w takich projektach

- Udział: Swiss System Biology: **SystemX** www.systemsx.ch
- Udział: European Consortium: **Endotrack** www.endotrack.org
- Udział: European Consortium: **Mitochcek** www.mitochcek.org
- Udział: European Consortium: **RIGHT** www.right-ip.org

5.5. Członkostwo w stowarzyszeniach

- Member of NVIDIA, CUDA Dresden – GPU
- Member of German Imaging Society
- Member of the Pattern Recognition Society
- Member Virtual Screening Society
- Advisory Board: Global Discover and Development

5.6. Międzynarodowe oraz krajowe nagrody za działalność naukowobadawczą

- Przyznanie pozycji w ramach konkursu na dyrektora działu w Instytucie Fraunhofer, Dresden
- Przyznanie pozycji na Tenure-Track Profesor na Uniwersytecie Philippsa w Marburgu
- Uznanie za utworzenie programu doktoranckiego pomiędzy Szwajcarią i Indiami: “Indo-Swiss”
- Przyznanie funduszu SystemX na utworzenie Serwisu Analiz Komputerowych przy centrum mikroskopowym na ETH w Zurychu.
- Przyznanie funduszu “Center of Excellence in Life Science Informatics (CELSI)”, Heidelberg, Open Source Consortium
- Uznanie firmy IBM i Thermofisher dla młodych naukowców “Young Life Scientists in HCS/HCA, High Content Analysis” , San Francisco, 2007

5.7. Wygłoszenie referatów na międzynarodowych konferencjach

1. Welcome Trust Center, Dundee, UK, 06.2006 – zaproszenie na wygłoszenie wykładu
2. RIGHT Symposium, Dresden, Germany, 09.2006 – zaproszenie na wygłoszenie wykładu
3. IIMCB, Warsaw, Poland, 10.2006 – zaproszenie na wygłoszenie wykładu
4. ChemBioNet, Frankfurt, Germany 12.2006 – zaproszenie na wygłoszenie wykładu
5. University of Chicago, Chicago, US, 01.2007 – zaproszenie na wygłoszenie wykładu
6. HCS, San Francisco, US, January, 01.2007 – zaproszenie na wygłoszenie wykładu

7. Screening Europe, Barcelona, Spain, 02.2007 – zaproszenie na wygłoszenie wykładu
8. CNIO, Madrid, Spain, 05.2007 – zaproszenie na wygłoszenie wykładu
9. HCA Europe, Berlin, Germany, 03.2008 – zaproszenie na wygłoszenie wykładu
10. SBS, St. Louis, US, 04.2008, Society for Biomolecular Sciences – workshop instruktor
11. SBS, St. Louis, US, 04.2008, Society for Biomolecular Sciences – sesja plenarna
12. Cell Based Assays, Munich, Germany, 05.2008, Informa – Biostatistics/ sesja plenarna
13. Protein Congress, Berlin, Germany, 10.2008 – zaproszenie na wygłoszenie wykładu
14. HCS, San Francisco, US, January, 01.2009 - workshop instruktor
15. HCS, San Francisco, US, January, 01.2009 - zaproszenie na wygłoszenie wykładu
16. High Content Analysis Conference, Barcelona, Spain, 03.2009, Informa – prowadzący konferencje, Biostatics
17. Global Discovery & Development, London, 03.2009 - zaproszenie na wygłoszenie wykładu na głównej sesji plenarnej
18. Translational Imaging in Drug Development, London, 04.2009 – zaproszenie na wygłoszenie wykładu
19. HCS, Boston, US, September, 09.2009 - zaproszenie na wygłoszenie wykładu
20. HCS, San Francisco, US, January, 01.2010 - zaproszenie na wygłoszenie wykładu
21. HCS, San Francisco, US, January, 01.2011 - - zaproszenie na wygłoszenie wykładu na głównej sesji plenarnej
22. SIG-KNIME, Zurich, March, 03.2011 – organizowanie konferencji
23. Screening Europe SBS, Hamburg, 06.2011 - zaproszenie na wygłoszenie wykładu na głównej sesji plenarnej
24. Screening Asia, Singapore, December, 11.2011 - sesja plenarna

5.8 Udział w międzynarodowych konferencjach ze względu na obowiązki administracyjne

1. HCS, San Francisco, US, January, 01.2009 - workshop instruktor
2. HCS, San Francisco, US, January, 01.2011 - workshop instruktor
3. HCS, San Francisco, US, January, 01.2012 - workshop instruktor
4. HCS, San Francisco, US, January, 01.2010 - workshop instruktor
5. HCS, Boston, US, September, 09.2009 - workshop instruktor
6. Russell Publishing - European Pharmaceutical Review, Dublin, June 2010- workshop instruktor

5.9. Zaprezentowanie posterów na międzynarodowych konferencjach

1. Kozak K., Kozak M., Korn K., Stoeter M., Fava E., Wagner J., Krausz E.: Screening Image Browser: Database and Tool for the Integration, Browsing and Analysis of Large Scale Image High-Content Screening Data. Screening Europe, Barcelona 2007, Spain
2. Kozak K.: HTS Data Scanner: Visualization und Evaluation of Clusters for Exploratory Analysis of Multi-parametric Cell Based Assays, High-Content Analysis 2007, January 23-26, 2007, San Francisco, USA.
3. Kozak K.: Storage and Statistical Practice in High-Content Screening Data Management, CehmBioNet, December 2006, Frankfurt, Germany.

4. Hoepfner S., Keller J.F., Korn K., Kozak K., Fava E., Grabner H., Klemm J., Schmitt J.; Wagner J., Krausz E., Zerial M.: Genome-Wide High-Content Screening in the Model Organism *C. elegans*. 12th Annual SBS Conference and Exhibition, Society of Biomolecular Sciences, 2006, September 17-21, Seattle, USA.
5. Kozak K.: Bioinformatics solutions for comprehensive management and analysis of large scale biological data. Trilateral Workshop, 2006, November 16-17, Warsaw, Poland
6. Kozak K.: Storage and Statistical Practice in High-Content Screening Data Management. High Content Analysis 2006, November 6-8, Vienna, Austria
7. Kozak K., Heninger A.K., Poser I., Slabicki M., Buchholz F., Krausz E.: Library Production: a Database for the Enzymatic Production of Small Interfering RNA (siRNA)". Life Science Automation 2006, September 14-15, Rostock, Germany
8. Kozak K., Kozak M., Stapor K., Grabner H., Korn K., Krausz, E.: LIMS: a Platform for Comprehensive Management and Analysis of Screening Data. Life Science Automation 2006, September 14-15, Rostock, Germany
9. Kozak K., Wagner J., Fava E., Stapor K., Grabner H., Korn K., Krausz E.: SIB: screening image browser. RIGHT Symposium on RNA Interference and High-Content Screening. 2006, September 11-12
10. Fava E., Moebius C., Grabner H., Korn K., Poser I., Lohmann A., Wagner J., Kozak K., Zerial M., Krausz, E.: High-Content Screening Using a Human Kinase siRNA Library: Screening Design and Applications. High-Content Analysis 2008, January 25-28, San Francisco, USA.
11. Kozak K., Tschida U., Beier G., Dasari V., Schlüpmann M., Schier H.: Unicode: Problem oder Lösung am Beispiel EDOC Jahrbuch Format. CPT-Herbsttagung 2003, Leipzig, Germany

5.10 Uznania za osiągnięcia naukowe

- Stanowisko dyrektora działu w Instytucie Fraunhofera, Drezno
- Przynanie tytułu Tenure-Track Professor of Philips Marburg University
- ETH Zurich Research Grant (ETH/O-20656-10 to UK), Switzerland
- SystemX Grant na utworzenie Computational Analysis Service at Screening facility
- Grant: Center of Excellence in Life Science Informatics (CELSI), Heidelberg,
 - Open Source Consortium
- IBM and Thermofisher Award for Young Life Scientists in HCS/HCA, High Content Analysis , San Francisco, January 22, 2007

6. Informacje biometryczne

Ocena biblioteczna:

Liczba cytowań: 267 (ISI Web of Science 1945-2013 z dnia 31.10.2013 r.)

Współczynnik Hirscha wg bazy Web of Science wynosi: 7

Łączna punktacja: IF = 91,39

KBN/MNiSW = 291

7. Informacje dodatkowe

Oprócz działalności naukowej zajmuję się także dydaktyką i popularyzacją bioinformatyki wśród „biologów-doświadczalników”. Prowadzę regularne wykłady na Wydziale Medycznym, Uniwersytetu w Dreźnie oraz jestem zapraszany do organizowania 1 lub 2-dniowych kursów/warsztatów analizy danych w technologii mikroskopowej RNAi, adresowanych głównie dla starszych studentów oraz młodych jak i doświadczonych pracowników naukowych. Takie kursy, składają się zwykle z kilku godzin wykładów i/lub kilku godzin ćwiczeń. Dotychczas przeprowadziłem następujące kursy:

- Letniej Szkoły Mikroskopowej w Zurychu (July 2007)
- SBS, St. Louis, US, 04.2008, Society for Biomolecular Sciences
- HCS, San Francisco, US, January, 01.2009
- HCS, Boston, US, January, 09.2009
- Symposium: Computational Biology, ETH Zurich, 09.2009
- HCS, San Francisco, US, January, 01.2010
- Russell Publishing - European Pharmaceutical Review, Dublin, June 2010

W roku 2009 zostałem zaproszony przez rektora Politechniki Radomskiej, Prof. Mariana Sułka na forum dyskusyjne dotyczące budowy nowego kierunku biotechnologii na tamtejszej uczelni. Profesor, w ramach współpracy zaproponował mi stanowisko gościnnego wykładowcy. W semestrze zimowym 2010 zacząłem cykl gościnnych wykładów dotyczący zagadnień informatyki i bioinformatyki dla studentów trzeciego i czwartego roku. Na Politechnice Radomskiej pełnię także rolę osoby wspierającej proces budowy infrastruktury i tworzenia koncepcji nowego kierunku.

W roku 2010 opublikowałem jako edytor książkę „Large Scale Data Handling in Biology” (wydawca: BookBone, Denmark, ISBN 978-87-7681-555-4, I Edition), składającą się z 4 rozdziałów (z czego wszystkie są mojego autorstwa). Opisane są tam zarówno metody komputerowe służące do zarządzania danymi WAG jak i analiza przepływu danych w tej technologii. Celem książki było stworzenie przewodnika dla nowo powstających jednostek WAG oraz studentów, którzy są zainteresowani zastosowaniem bioinformatycznych narzędzi do analizy danych mikroskopowych w swojej praktyce badawczej.

Posiadam także pracę w ramach działalności recenzenckiej. Byłem recenzentem następujących książek:

1. Book: “See Spot Count- A Primer of High Content Screening and Analysis Techniques”, John Wiley and Sohn
2. Patent procedure: Software for Time lapse Experiments in Live Cell Imaging

Wraz z moim zespołem nadal koncentrujemy się na dalszym odkrywaniu mechanizmów działania RNAi i szukaniu nowych funkcji genów, które umożliwiają komunikację pomiędzy kompartmentami wewnątrzkomórkowymi (takimi jak rybosomy i jądro komórkowe) w celu przekazywania sygnałów regulujących wzrost, podział lub śmierć komórek.