Encyklopedie subsetů CLL – unikátní bioinformatický nástroj a databáze pro analýzu subsetů stereotypních B buněčných receptorů u CLL

Encyclopedia of CLL Subsets – a Unique Bioinformatics Tool and Database for Analysis of Subsets of Stereotypical B-Cell Receptors in CLL

Reigl T.[1], Stránská K.[1,2], Bystrý V.[1], Krejčí A.[1], Grioni A.[1], Pospíšilová Š.[1,2], Darzentas N.[1], Plevová K.[1,2]

[1] CEITEC – Středoevropský technologický institut, Masarykova univerzita, Brno
[2] Interní hematologická a onkologická klinika LF MU a FN Brno

Souhrn

Východiska: Chronická lymfocytární leukemie (CLL) se vyznačuje vysokou klinickou i biologickou variabilitou. Ta je úzce spjata s řadou buněčných a molekulárních znaků, mezi něž patří sekvenční motivy B buněčných receptorů. Tyto motivy se u třetiny pacientů s CLL vyskytují v téměř identické (stereotypní) podobě, což je umožňuje zařadit do homogenních skupin, tzv. stereotypních subsetů CLL. Homogenita stereotypních subsetů není určena pouze sekvenčními motivy B buněčných receptorů, ale odráží se i v klinicko-biologických charakteristikách onemocnění. Pro zjednodušení přístupu k informacím o jednotlivých subsetech byl vytvořen veřejně dostupný webový nástroj Encyclopedia of CLL Subsets. ***Materiál a metody:*** Encyklopedie subsetů CLL vznikla jako součást bioinformatické platformy Antigen Receptor Research Tool (ARResT) vyvinuté speciálně pro analýzu, shlukování a anotaci imunoglobulinových sekvencí. Datový soubor od 7 500 CLL pacientů, na kterém je systém Encyklopedie postaven, pochází z mezinárodní studie Agathangelidis et al publikované v roce 2012 [1]. Analýzou tohoto souboru vznikl přehled hlavních stereotypních subsetů a jejich charakteristik. Další související klinické a cytogenomické informace o jednotlivých subsetech byly získány strojovým zpracováním dostupné literatury ze serveru PubMed a jsou pravidelně doplňovány a aktualizovány. ***Výsledky:*** Vytvořili jsme unikátní webovou aplikaci Encyklopedie subsetů CLL dostupnou na http://arrest.tools/subsets. Ta umožňuje interaktivní přístup k informacím o stereotypních subsetech CLL. Uživatel může získat a porovnávat základní informace o hlavních subsetech, vč. jejich klinických a cytogenomických vlastností. Tyto informace byly ručně zkontrolovány a vybrány ze strojově zpracovaných výsledků z databáze PubMed za pomoci odborníků v oblasti výzkumu CLL. V rámci Encyklopedie mají uživatelé také možnost přímo použít publikovaný nástroj ARResT/AssignSubsets a přiřadit vlastní sekvence B buněčných receptorů do hlavních subsetů. ***Závěr:*** Encyklopedie subsetů CLL je veřejně dostupný online nástroj, který usnadňuje přístup k nejnovějším poznatkům z oblasti výzkumu stereotypních subsetů u CLL a navíc umožňuje analýzu vlastních dat a interpretaci získaných výsledků. To má velký potenciál pro využití Encyklopedie v běžné praxi.

Klíčová slova

chronická lymfocytární leukemie – bioinformatika – imunogenetika – imunoglobuliny – stereotypní BCR

Podpořeno z programového projektu Ministerstva zdravotnictví ČR s reg. č. 16-34272A. Všechna práva vyhrazena.

This work was supported by Czech Ministry of Health grant No. 34272A. All rights reserved.

Autoři deklarují, že v souvislosti s předmětem studie nemají žádné komerční zájmy.

The authors declare they have no potential conflicts of interest concerning drugs, products, or services used in the study.

Redakční rada potvrzuje, že rukopis práce splnil ICMJE kritéria pro publikace zasílané do biomedicínských časopisů.

The Editorial Board declares that the manuscript met the ICMJE recommendation for biomedical papers.

Mgr. Tomáš Reigl
CEITEC – Středoevropský technologický institut
Masarykova univerzita
Kamenice 753/5
625 00 Brno
e-mail: tomas.reigl@gmail.com

Obdrženo/Submitted: 1. 3. 2019
Přijato/Accepted: 4. 3. 2019

Summary

Background: Chronic lymphocytic leukemia (CLL) is clinically and biologically highly variable disease which is closely related with multiple cellular and molecular markers, including sequence motifs of B-cell receptors. These motifs are highly similar (stereotyped) within one third of CLL patients and create homogeneous groups called stereotyped CLL subsets. The homogeneity is reflected also in clinical and biological characteristics of the disease. To facilitate access to the information about individual subsets, we have created a publicly available web-based tool Encyclopedia of CLL Subsets. ***Materials and methods:*** The Encyclopedia of CLL subsets belongs to our bioinformatics platform Antigen Receptor Research Tool (ARResT) developed for analysis, clustering, and annotation of immunoglobulin sequences. To gather primary knowledge about the subsets, we have analyzed a dataset of 7,500 CLL patients published by Agathangelidis et al in 2012 [1]. We have created an overview of major stereotyped subsets and their characteristics. Additional clinical and cytogenomic information about individual subsets has been obtained by machine text processing of available literature from server PubMed and is regularly updated. ***Results:*** We have created a unique web-based application Encyclopedia of CLL Subsets available from http://arrest.tools/subsets for an interactive access to the information about stereotyped CLL subsets. A user can obtain and compare basic information about the major subsets including their clinical and cytogenomic characteristics. These have been manually curated from machine processed results from PubMed database by experts in CLL research. Through the Encyclopedia's user interface, user can also directly use our published tool ARResT/AssignSubsets to assign new immunoglobulin sequences to the major subsets. ***Conclusion:*** The Encyclopedia of CLL Subsets is a publicly available online tool facilitating access to the most recent research knowledge about stereotyped CLL subsets and enabling analysis of own data and interpretation of the results. This gives the Encyclopedia a great potential for its use in clinical routine.

Key words

chronic lymphocytic leukemia – bioinformatics – immunogenetics – immunoglobulins – stereotyped BCR

Východiska

Chronická lymfocytární leukemie (CLL) se vyznačuje vysokou klinickou i biologickou variabilitou. Tato variabilita je úzce spjata s řadou buněčných a molekulárních znaků, mezi něž patří sekvenční motivy B buněčných receptorů (B-cell receptor – BCR). Tyto motivy se u třetiny pacientů s CLL vyskytují v téměř identické (stereotypní) podobě. Na jejich základě je možné pacienty zařadit do homogenních skupin, tzv. stereotypních subsetů CLL. Homogenita stereotypních subsetů není dána pouze sekvenčními motivy BCR, ale odráží se i v klinicko-biologických charakteristikách onemocnění. Pro zjednodušení přístupu k informacím o jednotlivých subsetech byl vytvořen veřejně dostupný webový nástroj Encyclopedia of CLL Subsets.

Metody

Encyklopedie subsetů CLL vznikla jako jedna ze součástí bioinformatické platformy Antigen Receptor Research Tool (ARResT) vyvinuté speciálně pro analýzu, shlukování a anotaci imunoglobulinových sekvencí.

Datový soubor, na kterém je celý systém Encyklopedie postaven, byl převzat z mezinárodního souboru přibližně 7 500 CLL pacientů publikovaného v roce 2012 [1]. Analýzou tohoto souboru jsme získali přehled o hlavních stereotypních subsetech (major sub-

Tab. 1. Charakteristiky subsetů a další klinické a cytogenomické informace.

Subset	Základní informace					
	počet pacientů	mutační status	klan *IGHV* genu	délka CDR3	*IGHV* gen	*IGHJ* gen
#1	183	nemutovaný	klan I	13	klan I	*IGHJ*4
#2	212	smíšený	klan III	9	*IGHV*3-21	*IGHJ*6
#3	42	nemutovaný	klan I	22	*IGHV*1-69	*IGHJ*6
#4	74	mutovaný	klan II	20	*IGHV*4-34	*IGHJ*6
#5	51	nemutovaný	klan I	20	*IGHV*1-69	*IGHJ*6
#6	68	nemutovaný	klan I	21	*IGHV*1-69	*IGHJ*3
#7H	23	nemutovaný	klan I	24	*IGHV*1-69	*IGHJ*6
#8	35	nemutovaný	klan II	19	*IGHV*4-39	*IGHJ*5
#12	22	nemutovaný	klan I	19	*IGHV*1-2	*IGHJ*4
#14	21	mutovaný	klan II	10	*IGHV*4-4	*IGHJ*4
#16	26	mutovaný	klan II	24	*IGHV*4-34	*IGHJ*6
#28A	23	nemutovaný	klan I	17	*IGHV*1-2	*IGHJ*6
#31	23	nemutovaný	klan III	21	klan III	*IGHJ*6
#59	22	nemutovaný	klan I	12	klan I	*IGHJ*4
#64B	20	nemutovaný	klan III	21	*IGHV*3-48	*IGHJ*6
#77	25	mutovaný	klan II	14	*IGHV*4-59	*IGHJ*4
#99	20	nemutovaný	klan I	14	klan I	*IGHJ*4
#201	23	mutovaný	klan II	17	*IGHV*4-34	*IGHJ*3
#202	16	nemutovaný	klan III	14	klan III	*IGHJ*4

mut – mutace

Tab. 1 – pokračování. Charakteristiky subsetů a další klinické a cytogenomické informace.

Subset	Klinické informace					Cytogenomické informace	
	věk při diagnóze	průběh nemoci	čas do první terapie	rizika	zdroje	mutace	zdroje
#1	–	agresivní	1,6 (0–8)	zvýšené riziko imunitní trombocytopenie	[3,4]	TP53 mut 16 % (21/135 případů) NFKBIE mut 15 % (17/112 případů) NOTCH1 mut 19–27 % del(13q): 20% (18/89 případů)	[8–10]
#2	–	agresivní	1,9 (0–7,9)	–	[3]	SF3B1 mut 36–45 % ATM mut 26 % (21/81 případů) TP53 aberace TP53 mut 0–5 %, del(17p) 0–4 %	[7,8,10–13]
#3	–	agresivní	2,7 (0–5,5)	zvýšené riziko autoimunní hemolytické anémie	[3,5]	SF3B1 mut 46 % (12/26 případů)	[8]
#4	mladí pacienti	velmi indolentní	11,0 (0,9–13,5)	–	[3,6]	–	–
#5	–	agresivní	1,8 (0–5,5)	–	[3]	–	–
#6	–	agresivní	1,6 (0–5,8)	–	[3]	NOTCH1 mut 22 % (10/45 případů)	[8]
#7H	–	agresivní	2,8 (0–7,2)	zvýšené riziko imunitní trombocytopenie	[3,4]	TP53 defekty 26 % (12/46 případů)	[13]
#8	–	agresivní	1,5 (0–8,1)	vysoké riziko transformace do Richterova syndromu	[3,7]	trizomie 12 60 % (15/25 případů) NOTCH1 mut 30 % (13/43 případů)	[3,8]
#12	–	–	–	–	–	–	–
#14	–	–	–	–	–	–	–
#16	mladí pacienti	velmi indolentní	–	–	[3,6]	–	–
#28A	–	–	–	–	–	–	–
#31	mladí pacienti	agresivní	1,0 (0–5,2)	–	[3]	–	–
#59	starší pacienti	agresivní	1,0 (0–4,7)	–	[3]	NOTCH1 mut 33 % (6/18 případů)	[8]
#64B	–	–	–	–	–	–	–
#77	mladí pacienti	indolentní	9,2 (0,2–13,0)	–	[3]	del(13q) 78 % (14/18 případů)	[3]
#99	–	–	–	–	–	TP53 mut 33 % (6/18 případů)	[8]
#201	–	indolentní	6,8 (0,2–10,3)	–	[3,6]	–	–
#202	–	–	–	–	–	–	

mut – mutace

sets) a o jejich charakteristikách. Nejdůležitější charakteristiky jsou v aplikaci zobrazeny pro všechny hlavní subsety – velikost subsetu (tj. počet pacientů), mutační status variabilní oblasti genu pro těžký řetězec imunoglobulinu (IGHV), délka CDR3 oblasti imunoglobulinové přestavby, klan *IGHV* genu, samotný *IGHV* a *IGHJ* gen charakteristický pro daný subset a grafické zobrazení („logo") sekvence CDR3 oblasti. Další související klinické a cytogenomické informace o jednotlivých stereotypních subsetech byly získány zpracováním dostupné literatury.

Publikace využité pro vytvoření přehledu klinicko-biologických charakteristik hlavních subsetů byly získány strojovou metodou pro zpracování textu ze serveru PubMed a jsou pravidelně doplňovány a aktualizovány. V nejnovější verzi Encyklopedie byly publikace prostudovány zkušenými odborníky a byly z nich vybrány důležité informace, které jsou prezentovány společně s odkazem na příslušnou literaturu.

V rámci Encyklopedie subsetů CLL mají uživatelé rovněž možnost přímo použít publikovaný nástroj ARResT/AssignSubsets [2] a přiřadit vlastní sekvence BCR do hlavních stereotypních subsetů.

Výsledky

Vytvořili jsme unikátní webovou aplikaci Encyklopedie subsetů CLL veřejně dostupnou na http://arrest.tools/subsets. Ta umožňuje interaktivní přístup k informacím o stereotypních subsetech CLL. Díky souhrnnému přehledu jednotlivých hlavních subsetů může uživatel získat a porovnávat základní informace. Důležitou součástí přehledu jsou zejména klinické a cytogenomické vlastnosti. Tyto informace byly ručně extrahovány ze strojově zpracovaných výsledků databáze PubMed za pomoci expertů v oblasti výzkumu CLL. Charakteristické klinické vlastnosti zahrnují např. věk při diagnóze, pravděpodobný klinický průběh nemoci, čas do první terapie a závažná rizika spojená s daným subsetem. Uživatel se dále dozví informace o mutacích typických pro daný subset vč. jejich frekvence a počtu v subsetu, které byly zjištěny v publikovaných kohortách (tab. 1).

Díky automatizovanému přístupu k veřejně dostupným elektronickým publikacím na serveru PubMed mají uživatelé k dispozici nejnovější poznatky o stereotypních CLL subsetech v přehledné podobě. Zpracované publikace jsou prezentovány v tabulce, která obsahuje konkrétní zmínky o jednotlivých stereotypních subsetech v plném textu publikace, umístění zmínky v plném textu, rok zveřejnění a identifikační číslo serveru PubMed (PMID) s odkazem na dostupné elektronické zdroje. V této tabulce je použito barevné kódování pro jednotlivé stereotypní subsety i pro relevantní klíčová slova (např. názvy genů, závažnost onemocnění atd.) pro snadnější orientaci a pochopení extrahované informace.

Od roku 2006 bylo publikováno více než 500 článků zmiňujících se o stereotypních CLL subsetech, z nich bylo manuálně vybráno stěžejních 236, které poskytují nová pozorování nebo ucelený přehled klinicko-biologických vlastností subsetů. Výskyt informací o stereotypních subsetech v publikacích má zjevnou rostoucí tendenci. Nejčastěji se vyskytují informace o subsetech #1, #2 a #4 – od roku 2016 je každý z nich uveden ve více než 60 publikacích.

Diskuze

Encyklopedie CLL subsetů byla vytvořena v reakci na zvyšující se význam stereotypních subsetů a jejich častější využití ve výzkumu CLL a také v klinické praxi. Tento unikátní nástroj a databáze poskytují ucelené informace v dostupné a snadno zpracovatelné formě. S rostoucím množstvím publikací s biomedicínskou tematikou může snadno dojít k přehlcení čtenáře nepodstatnými informacemi a k přehlédnutí podstatného. Důležitým faktorem udržitelnosti je pravidelná aktualizace dostupné literatury, na jejíž automatizaci v současné chvíli pracujeme. Díky Encyklopedii subsetů CLL se dostane uživatel k seznamu veškeré aktuální klíčové literatury a bližším informacím o konkrétních subsetech snadno a efektivně.

Závěr

Rostoucí význam stereotypních subsetů CLL se odráží v množství publikací pojednávajících o tomto tématu. Pro usnadnění přístupu k adekvátní literatuře a zasazení publikovaných dat do širších souvislostí vznikla Encyklopedie subsetů CLL. Jedná se o veřejně dostupný online nástroj zpřístupňující nejnovější poznatky z dané oblasti a umožňující analýzu vlastních dat a interpretaci výsledků v kontextu konkrétního subsetu i ostatních subsetů. Námi vyvinutý přístup umožňuje širší pohled na danou problematiku a má velký potenciál pro využití v běžné praxi.

Literatura

1. Agathangelidis A, Darzentas N, Hadzidimitriou A et al. Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. Blood 2012; 119(19): 4467–4475. doi: 10.1182/blood-2011-11-393694.
2. Bystry V, Agathangelidis A, Bikos V et al. ARResT/AssignSubsets: a novel application for robust subclassification of chronic lymphocytic leukemia based on B cell receptor IG stereotypy. Bioinformatics 2015; 31(23): 3844–3846. doi: 10.1093/bioinformatics/btv456.
3. Baliakas P, Hadzidimitriou A, Sutton LA et al. Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicentre study. Lancet Haematol 2014; 1(2): e74–e84. doi: 10.1016/S2352-3026(14)00005-2.
4. Visco C, Maura F, Tuana G et al. Immune thrombocytopenia in patients with chronic lymphocytic leukemia is associated with stereotyped B-cell receptors. Clin Cancer Res 2012; 18(7): 1870–1878. doi: 10.1158/1078-0432.CCR-11-3019.
5. Maura F, Visco C, Falisi E et al. B-cell receptor configuration and adverse cytogenetics are associated with autoimmune hemolytic anemia in chronic lymphocytic leukemia. Am J Hematol 2013; 88(1): 32–36. doi: 10.1002/ajh.23342.
6. Xochelli A, Baliakas P, Kavakiotis I et al. Chronic lymphocytic leukemia with mutated IGHV4-34 receptors: shared and distinct immunogenetic features and clinical outcomes. Clin Cancer Res 2017; 23(17): 5292–5301. doi: 10.1158/1078-0432.CCR-16-3100.
7. Rossi D, Spina V, Cerri M et al. Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome. Clin Cancer Res 2009; 15(13): 4415–4422. doi: 10.1158/1078-0432.CCR-08-3266.
8. Sutton LA, Young E, Baliakas, P et al. Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors. Haematologica 2016; 101(8): 959–967. doi: 10.3324/haematol.2016.141812.
9. Mansouri L, Sutton LA, Ljungström V et al. Functional loss of IκBε leads to NF-κB deregulation in aggressive chronic lymphocytic leukemia. J Exp Med 2015; 212(6): 833–843. doi: 10.1084/jem.20142009.
10. Strefford JC, Sutton LA, Baliakas P et al. Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset# 2. Leukemia 2013; 27(11): 2196–2199. doi: 10.1038/leu.2013.98.
11. Jeromin S, Haferlach C, Dicker F et al. Differences in prognosis of stereotyped IGHV3-21 chronic lymphocytic leukaemia according to additional molecular and cytogenetic aberrations. Leukemia 2016; 30(11): 2251–2253. doi: 10.1038/leu.2016.189.
12. Navrkalova V, Young E, Baliakas P et al. ATM mutations in major stereotyped subsets of chronic lymphocytic leukemia: enrichment in subset# 2 is associated with markedly short telomeres. Haematologica 2016; 101(9): e369–e373. doi: 10.3324/haematol.2016.142968.
13. Malcikova J, Stalika E, Davis Z et al. The frequency of TP 53 gene defects differs between chronic lymphocytic leukaemia subgroups harbouring distinct antigen receptors. Br J Haematol 2014; 166(4): 621–625. doi: 10.1111/bjh.12893.