

VYUŽITÍ DVOUROZMĚRNÉHO PŘÍMÉHO AGLOMERATIVNÍHO SHLUKOVÁNÍ PRO ANALÝZU IMUNITNÍ ODPOVĚDI REKOMBINANTNÍCH KONGENNÍCH KMENŮ

IGOR KÁRNÍK A MARIE LIPOLDOVÁ

ABSTRAKT. Various components of the immune response are controlled by multiple genes. We have used a special tool for genetic analysis of these multigenically-controlled biological traits: the Recombinant Congenic Strains (RCS). The series of RCS comprises homozygous mouse strains. Each RC strain of OcB / Dem Series contains a different, random, set of about 12.587.5 used at least 10 mice from each strain (total number 200 č 300 mice). One task of the statistical analysis was to establish whether RC strains statistically differ in immune response to different factors (e.g. alloantigens or infection by *Leishmania major*). We have used mainly MANOVA with mixed effects, in some cases we have also used MANCOVA. Another studied question was to find similarity between individual Recombinant Congenic Strains. We have used agglomerative methods of hierarchical cluster analysis (mainly complete linkage and Ward method) and the direct two-way cluster analysis. This paper describes this interesting method and gives examples of some practical applications. This work was supported by grants NM/28-3 and OK 394 from the Ministry of Health and the Ministry of Education of the Czech Republic, respectively. We have used programmes STATGRAPHICS PLUS 7, SPSS 8.01, and STATISTICA 4.3 (for two-way direct cluster).

Абстракт. В этой статье изучается кластер анализ для статистического контроля в генетике.

1. VĚCNÝ ÚVOD DO PROBLEMATIKY

Většina důležitých civilizačních chorob (hypertenze, diabet, alergie, duševní onemocnění) jsou polygenně kontrolovány, to znamená, že jejich vznik je ovlivněn mnoha různými geny. K jejich studiu používáme nový genetický model, speciálně připravený pro analýzu polygenně kontrolovaných biologických znaků, rekombinantní kongenní kmeny (RCS) (Demant and Hart, 1986). V tomto systému je křížením dvou rodičovských kmenů připravena série 20 homozygotních RCS. Jeden rodičovský kmen slouží jako kmen základní, druhý je dárce genetického materiálu. Každý z 20 různých RCS nese asi 12.5dárce a 87.5geny dárce, které se podílejí na multigenní kontrole vnímavosti k onemocnění, rozděleny mezi různé kmeny, kde mohou být studovány odděleně. Systém RCS tak převádí rozdíl v mnoha genech mezi dárce a základním kmenem na řadu rozdílů v jednom genu (nebo několika málo genech) mezi jednotlivými RCS a kmenem základním. Odlišnosti ve funkci jednoho genu pak mohou být předmětem vazebné analýzy a funkčních studií. Jednotlivé RCS jsou analogické s tzv. izolovanými populacemi, známými z lidských studií vazebného disekvilibria (Frankel and Schork, 1996). Tento systém je zvláště vhodný ke studiu infekcí, jak jsme ukázali na příkladu studia vnímavosti k parazitu *Leishmania major* (Demant et

2000 *Mathematics Subject Classification*. Primary 62P20.

Klíčová slova. Shluková analýza.

Práce na tomto článku byla financována z grantů IGA MZ ČR NM/28 č. 3 a MŠMT OK 394.

al., 1996; Lipoldová et al., 2000). Dále jsme sledovali produkci cytokinů, proliferační odpověď stimulovanou cytokiny, mitogeny a antigeny.

Jako příklad použití dvourozměrné shlukové analýzy jsme vybrali jeden z našich experimentů, v nichž byla sledována odpověď rekombinantních kongenních kmenů série OcB/Dem k lymfocytům myších kmenů DBA/1, CBA, BALB/c a C57BL/10. Každý kmen této série nese různou kombinaci 12.5genů kmene B10.O20 a 87.5analyzovali odpověď minimálně 10 myší. Rodičovské kmeny O20 a B10.O20 se v odpovědi k těmto antigenům neliší, zatímco RCS vykazovaly značnou heterogenitu odpovědi (Havelková et al., připravovaný manuskript).

2. STATISTICKÁ ANALÝZA SE ZVLÁŠTNÍM PŘIHLÉDNUTÍM K VYUŽITÍ DVOUROZMĚRNÉHO PŘÍMÉHO CLUSTEROVÉ ANALÝZY

Úvodem poznamenejme, že do každého experimentu vstupovalo 200 č 300 myší (v námi uváděném případě jich bylo konkrétně 230). Hlavním úkolem statistické analýzy pak bylo vytipovat jednotlivé faktory, resp. jejich interakce, statisticky významně působící na odlišnosti mezi jednotlivými kmeny. Pro tuto úlohu byla použita MANOVA se smíšenými efekty (některých případech i MANCOVA). Druhým hlavním úkolem statistické analýzy pak bylo určení podobnosti mezi jednotlivými kmeny. Pro tento účel byly použity aglomerativní metody hierarchické cluster analýzy (z interpretačního hlediska se pak jako nejvhodnější ukázala metoda nejbližšího souseda a metoda Wardova). Jako reprezentanty proměnných u jednotlivých kmenů jsme převážně uvažovali aritmetické průměry těchto proměnných (v některých případech i jejich mediány).

Pro clusterovou analýzu jsme používali program STATISTICA a při této příležitosti jsme se setkali s dvourozměrnou shlukovou analýzou (v originále Two č way Joining), která je popsána (pokud jsme zjistili) pouze v publikaci Hartigan (1975). Zvláštěností této metody je fakt, že vytváří dvourozměrné shluky mezi řádky a sloupci datové matice. V citované práci jsou uvažovány jak divizní, tak i aglomerativní varianty této metody (autor však používá poněkud jiné terminologie). Program STATISTICA pak používá pouze aglomerativní metodu. S velmi podrobným popisem algoritmů jednotlivých metod a programů ve FORTRANu se čtenář může seznámit v citované práci na str. 251 č 298.

Uvedme však, že citovaná publikace byla vydána v roce 1975 a není orientována na grafický výstup, zatímco z interpretačního hlediska je zajímavý právě grafický výstup poskytovaný paketem STATISTICA.

Konkrétní interpretaci si ukážeme na námi uvedeném příkladě. Disponujeme výstupy jak pro všechny uvažované kmeny, tak i bez kmene OcB - 3. Interpretace je pro druhý případ. Grafický výstup je uveden v příloze. Datová matice byla uspořádána tak, že ve sloupcích byly uvedeny jednotlivé kmeny a v řádcích aritmetické průměry jednotlivých proměnných. V grafu pak jsou na ose x uvedeny čísla jednotlivých kmenů a na ose y jednotlivé proměnné. Uvažujme např. proměnnou BALB/c. Je zřejmé, že z hlediska této proměnné tvoří 1. shluk kmeny O20, OcB -19, OcB - 8, OcB - 1, OcB - 15, OcB - 2, OcB - 10, OcB - 12, OcB - 7, B10.O20. Druhý shluk je tvořen kmeny OcB - 21, OcB - 5, OcB - 14 a OcB - 11. Kmeny OcB-16 a OcB - 17 jsou sloučeny do třetího clusteru, zatímco kmeny OcB - 6 a OcB -9 tvoří samostatné shluky (tedy čtvrtý a pátý). Analogická interpretace by byla i u dalších dvourozměrných clusterů. Závěrem uvedme, že pořadí kmenů na ose x odpovídá přibližně shlukům hierarchického aglomerativního clustrování podle metody

nejvzdálenějšího souseda, či příbuzné Wardovy metody. Pro dobrou interpretaci je tedy vhodné použít hierarchického i dvourozměrného shlukování.

Tato metoda nám může podat první informaci o genetické podobnosti zkoumaných kmenů a na jejím základě můžeme lépe vybrat kmeny pro další studium. Protože kmen OcB č 9 od všech ostatních nejvíce lišil, použili jsme jej pro genetickou analýzu a identifikovali jsme lokus Alan2, který kontroluje odpověď aloantigenům C57BL/10, BALB/c a CBA (Havelková et al., v tisku).



3. ZÁVĚR

Samotným závěrem pouze poznamenejme, že přes zcela evidentní přednosti a značnou aplikabilitu, má námi předkládaná metoda i slabiny. Za největší z nich považujeme to, že de facto chybí její přesný matematický popis. Pokud je nám známo je totiž uvedena v námi citované práci Hartigan (1975), která je napsána ryze programátorským stylem. Obsahuje pouze sáhodlouhé algoritmy a programy. V Hartiganem citované literatuře k této metodě jsou uvedeny pouze odkazy na čtyři FORTRANovské subrutiny. Dále je nutno si uvědomit, že daná publikace byla vydána již v roce 1975, tedy v období velkých sálových počítačů, kdy byl rozvoj počítačové grafiky na samém počátku a neobsahuje byť i jen zmínku o grafickém výstupu zmiňované metody (viz Příloha), který je na ní beze sporu nejzajímavější. Úplným závěrem konstatujeme, že podobnou problematikou se zabývají např. práce Lefkovitch (1987), Bezdek (1987), Legendre (1987) a Pleszynska et al (1995) aj.

Příloha

Kmen 0 přitom značí kmen O20.

Kmen 22 přitom značí kmen B10.O20.

Ostatní kmeny pak OcB č Dem (kmen 19 jest identický se kmenem OcB č 19).

Pomocný slovníček

Leishmania major = parazitický prvok.

Cytokin = (růstový faktor) bílkoviny produkované buňkami, které působí na další buňky.

Mitogen = látka stimulující množení buněk.

Antigen = látka schopná vyzvat imunitní odpověď.

Proliferace = buněčný růst kdy dochází ke zmnožení buněk.

Lymfocyt = typ bílých krvinek.

BALB/c, DBA, CBA, C57BL/10 = názvy myších kmenů.

Alan 2 (Alloantigen response 2) = gen, který kontroluje odpověď k aloantigenům (antigeny stejného živočišného druhu).

Literatura

- 1 Atkinson A. C. (1989). Plots, Transformations and Regression. Oxford .
- Bezdek, J. C (1987). Some non-standard Clustering Algorithms. NATO Asi Series, Vol. G14. Springer Verlag Berlin. 225 č 287
- 2 Demant P. and Hart A.A.M. (1986). Recombinat congenic strains: a new tool for analyzing genetic traits determined by more than one gene. Immunogenetics 24: 416-422.
- 3 Demant P., Lipoldová M., Svobodová M. (1996). Resistance to Leishmania major in mice (Technical Comment). Science 274: 1392.
- 4 Frankel W.N., Shork N.J (1996) Whos afraid of epistasis? Nature Genet. 14: 371-373.
- 5 Havelková H., Demant P., Lipoldová M. A new type of genetic regulation of allloresponse. A novel locus on mouse chromosome 4, Alan 2 controls MLC reactivity to three different alloantigens: C57BL/10, BALB/C and CBA. Genes and Immunity, 1: in press 2000.
- 6 Havelková H., Holář V., Kárník I., Demant P., Lipoldová M.: Separation of multiple non-MHC genes controlling proliferative response to alloantigens using recombinant congenic strains (připravovaný manuscript).
- 7 Hartigan J.A. (1975) Clusterig algorithms. Wiley and Sons, New York.
- 8 Hebák P., Hustopecký J., (1987): Vícerozměrné statistické metody s aplikacemi. SNTL/Alfa. Praha..
- 9 Lefkovich L.P. (1987). Species Associations and condicional clustering: clustering with or without pairwise resemblances. NATO ASI Series, Vol. G14. Springer Verlag Berlin. 309- 331.
- 10 Legendre P. (1987). Constrained Clustering. NATO ASI Series, Vol. G14. Verlag Berlin. 289 č 307.
- 11 Lipoldová M., Svobodová M., Krulová M., Havelková H., Badalová J., Nohýnková E., Holář V., Hart A.A.M., Volf P., Demant P. Susceptibility to Leishmania major infection in mice: multiple loci and heterogeneity of immunopathological phenotypes. Genes and Immunity 1: 200-206,2000.
- 12 Pleszczyńska E., Kowalczyk, T., Szczepny W., Cionk A. (1995). New tools Statistical Data Analysis. Biocybernetics and Biomedical engineering 15/95.