

به کارگیری روش‌های خوشه‌بندی در ریزآرایه DNA

حمید علوی مجد*، محسن واحدی**، یدالله محرابی***، بهار نقوی****

* دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی
** گروه آمار زیستی، دانشگاه علوم پزشکی شهید بهشتی
*** دانشکده بهداشت، دانشگاه علوم پزشکی شهید بهشتی
**** دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی شهید بهشتی

چکیده

سابقه و هدف: به کارگیری فناوری ریزآرایه DNA که امکان بررسی بیان هزاران ژن را به طور همزمان در حداقل زمان ممکن می‌سازد، در سال‌های اخیر موجب تولید حجم انبوهی از داده‌های بیان ژنی شده است. تحلیل آماری این داده‌ها شامل مواردی چون نرمال‌سازی، خوشه‌بندی، طبقه‌بندی است. هدف این مقاله بررسی نحوه به کارگیری روش‌های آماری خوشه‌بندی در داده‌های ریز آرایه DNA است.

روش بررسی: در این تحقیق داده‌های سرطان پستان وانتور و همکاران (۲۰۰۲) مربوط به جهش‌های ژنتیکی *BRCA1* و *BRCA2*، تحلیل شده است. مجموعه داده‌ها شامل ۱۸ بیمار با جهش *BRCA1* و ۲ بیمار با جهش *BRCA2* است. داده‌های بیان ژنی سرطان پستان با استفاده از روش‌های آماری سلسله مراتبی و غیر سلسله مراتبی خوشه‌بندی گردید. در هر دو روش خوشه‌بندی، داده‌ها به دو خوشه تقسیم شدند. روش‌های مختلف خوشه‌بندی با توجه به گروه‌بندی واقعی (*BRCA1*, *BRCA2*) مقایسه شدند. نرم‌افزار R برای تحلیل داده‌ها استفاده شد.

یافته‌ها: ویژگی روش خوشه‌بندی سلسله مراتبی در تشخیص ژن *BRCA1* ۹۴ درصد و حساسیت آن ۱۰۰ درصد بدست آمد. ویژگی روش خوشه‌بندی غیر سلسله مراتبی در تشخیص ژن *BRCA1* ۸۹ درصد و حساسیت آن ۱۰۰ درصد بدست آمد که نشان‌دهنده عملکرد مناسب دو روش خوشه‌بندی است. روش خوشه‌بندی سلسله مراتبی بر اساس ادغام بر حسب میانگین مناسب‌ترین روش در بین همه روش‌های بررسی شده است. نمونه شماره ۹۵ طبق نتایج همگی روش‌های خوشه‌بندی در گروه *BRCA2* قرار گرفت در صورتی‌که بر اساس یافته‌های بالینی در گروه *BRCA1* قرار دارد.

نتیجه‌گیری: با توجه به انطباق قابل توجه نتایج خوشه‌بندی با گروه‌بندی واقعی داده‌ها، می‌توان از این روش‌های آماری در مواردی که اطلاع دقیقی از گروه‌بندی واقعی داده‌ها در دست نیست، استفاده کرد. به علاوه نتایج خوشه‌بندی ممکن است زیر گروه‌هایی از نمونه‌ها را به نحوی متمایز کند که برای انطباق آن با یافته‌های بالینی، پژوهش‌های آزمایشگاهی یا بالینی جدیدی لازم باشد.

واژگان کلیدی: ریزآرایه DNA، خوشه‌بندی، بیان ژن، سرطان پستان

مقدمه

دومین علت مرگ و میر به‌شمار می‌رود. در کشورهای در حال توسعه نیز بیماری سرطان در ردیف مسایل مهم بهداشتی درمانی محسوب و روند آن رو به افزایش است (۱). سرطان پستان شایع‌ترین سرطان زنان در سراسر دنیا بوده و علت عمده مرگ‌های ناشی از سرطان در زنان است. محققین در مطالعات مختلف نقش عوامل متعدد خطر سازی را برای سرطان پستان شناسایی کرده‌اند. شناخت و کشف علت‌ها به جلوگیری از بیماری در افراد سالم و نسل‌های بعد کمک

در حال حاضر سرطان یکی از مسایل مهم و اصلی بهداشت و درمان در ایران و تمام دنیا می‌باشد. در آمریکا و تعدادی از کشورهای اروپایی، این بیماری بعد از بیماری‌های قلبی-عروقی

آدرس نویسنده مسئول: تهران، دانشگاه علوم پزشکی شهید بهشتی، گروه آمار زیستی دانشکده پیراپزشکی، دکتر حمید علوی مجد (email: alavimajd@gmail.com)

تاریخ دریافت مقاله: ۱۳۸۵/۷/۳۰

تاریخ پذیرش مقاله: ۱۳۸۶/۱/۳۱