تأثیر نتایج درمانی مرتبط با ویژگی‌های مولکولی کم‌خونی آپلاستیک کودکان[1]

فرزان رحمانی
مریم صادق‌آبادی
محمد رضا مزروعی

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف

۱۸ بهمن ۱۴۰۳

چکیده

آنمی آپلاستیک شدید با مغز استخوان کم‌سلولی و سیتوپنی محیطی مشخص می‌شود. MSCها نقش مهمی در توسعه سلول‌های بنیادی خون‌ساز (HSCs) و ایجاد یک محیط ریزمناسب برای خون‌سازی دارند. بررسی مولکولی مسیرهای نگهداری تلومر و پروفایل بیان ژنی MSCها می‌تواند برای مداخلات درمانی در بیماران کودکان مبتلا به آنمی آپلاستیک مهم باشد.

این مطالعه شامل ۱۰ بیمار کودک مبتلا به آنمی آپلاستیک و ۸ کودک سالم هم‌سن به عنوان گروه کنترل است. نمونه‌های خون محیطی از این افراد جمع‌آوری شد و طول تلومر لکوسیت‌ها با روش PCR زمان واقعی اندازه‌گیری شد. همچنین، داده‌های پروفایل بیان ژنی مبتنی بر ریزآرایه (GSE۳۳۸۱۲) از MSCهای پنج بیمار آنمی آپلاستیک کودکان، دانلود و تحلیل شد و با پنج فرد سالم به عنوان گروه کنترل مقایسه گردید.

طول تلومر در بیماران کودک مبتلا به آنمی آپلاستیک به طور قابل‌توجهی کوتاه‌تر از افراد سالم هم‌سن بود. جالب اینکه، یک زیرگروه شامل دو بیمار کودک دارای طول تلومر متوسطی بودند که مشابه گروه کنترل بود. بر اساس تحلیل مسیر بیان ژنی مرتبط با نگهداری تلومر، ژن TERF۲ به طور قابل‌توجهی در بیماران با تلومر کوتاه‌تر تنظیم منفی شده بود، اما در بیماران با تلومر متوسط تغییری مشاهده نشد. همچنین، پروفایل بیان ژنی MSCها نشان داد که سه ژن A۱۵, TMSB۶۷, MKI۳L۲GAS به طور افتراقی بیان شده و با نتایج درمانی مرتبط هستند.

الگوی طول تلومر و پروفایل بیان ژنی MSCها و مسیرهای نگهداری تلومر ممکن است به عنوان یک نشانگر زیستی بالقوه برای پیش‌بینی نتایج درمانی آنمی آپلاستیک کودکان مورد استفاده قرار گیرد.

کلمات کلیدی

آنمی آپلاستیک، کودکان، تلومر، سلول‌های بنیادی مزانشیمی، درمان

---

[1] Enhancing Modeling and Simulations with Virtual Reality and Augmented Reality

۱- مقدمه

کم‌خونی آپلاستیک (Aplastic Anemia) یک بیماری نادر و خطرناک است که در اثر آسیب به مغز استخوان ایجاد می‌شود. مغز استخوان که وظیفه تولید سلول‌های خونی را بر عهده دارد، در این بیماری دچار کاهش شدید فعالیت شده و به‌جای سلول‌های خونی طبیعی، سلول‌های بسیار کمی تولید می‌کند یا به‌طور کامل عملکرد خود را از دست می‌دهد. این وضعیت باعث کاهش شدید سه نوع اصلی سلول‌های خونی می‌شود: گلبول‌های قرمز، گلبول‌های سفید و پلاکت‌ها. به این کاهش همزمان سلول‌ها اصطلاحاً پان‌سیتوپنی گفته می‌شود.

علائم بیماری کم‌خونی آپلاستیک عمدتاً ناشی از کاهش تعداد سلول‌های خونی است و شامل خستگی شدید به دلیل کمبود گلبول‌های قرمز و کاهش توانایی حمل اکسیژن، کبودی و خونریزی آسان مانند خونریزی لثه یا بینی به علت کاهش پلاکت‌ها، و عفونت‌های مکرر یا طولانی‌مدت به دلیل کاهش گلبول‌های سفید می‌شود، اگرچه عفونت‌ها در مراحل اولیه بیماری کمتر مشاهده می‌شوند. علاوه بر این، علائم عمومی مانند رنگ‌پریدگی، ضعف عضلانی، و کاهش تحمل فعالیت‌های بدنی نیز در میان بیماران رایج است.

برای بیماران مشکوک به کم‌خونی آپلاستیک، تشخیص سریع از اهمیت بالایی برخوردار است، زیرا در صورت عدم درمان، این بیماری می‌تواند تهدیدکننده حیات باشد. ابزارهای تشخیصی شامل آزمایش‌های خون (برای بررسی میزان سلول‌های خونی) و نمونه‌برداری از مغز استخوان برای ارزیابی سطح سلول‌سازی هستند.

برای درمان کم‌خونی آپلاستیک دو روش اصلی وجود دارد: درمان‌های سرکوب‌کننده سیستم ایمنی (IST) و پیوند مغز استخوان (BMT). درمان‌های سرکوب‌کننده سیستم ایمنی برای بیمارانی به کار می‌رود که امکان پیوند مغز استخوان ندارند و هدف آن کاهش پاسخ‌های خودایمنی بدن است که به سلول‌های مغز استخوان آسیب می‌رسانند. در مقابل، پیوند مغز استخوان، به‌ویژه در بیماران جوان‌تر و کودکانی که اهداکننده مناسبی دارند، روشی بسیار مؤثر است که عملکرد طبیعی مغز استخوان را بازمی‌گرداند. علاوه بر این، مراقبت‌های حمایتی، مانند تزریق خون و پیشگیری از عفونت‌ها، در کاهش عوارض بیماری نقش مهمی ایفا می‌کند.

شیوع کم‌خونی آپلاستیک در سراسر جهان متفاوت است. بر اساس داده‌های "مطالعه بین‌المللی آگرانولوسیتوز و کم‌خونی آپلاستیک" (IAAAS)، این بیماری در آسیا شایع‌تر است، به طوری که میزان ابتلا در آسیا ۲ تا ۳ برابر بیشتر از کشورهای غربی مانند اروپا و آمریکای شمالی گزارش شده است. عوامل محیطی، مانند تماس با مواد شیمیایی خاص، مصرف داروهای خاص، و عفونت‌های ویروسی، می‌توانند در بروز بیماری نقش داشته باشند. علاوه بر این، عوامل ژنتیکی و تنوع‌های قومی نیز ممکن است تأثیرگذار باشند.

عوامل محیطی مانند تماس با مواد شیمیایی و عفونت‌ها، به‌علاوه عوامل ژنتیکی، ممکن است خطر ابتلا به بیماری را افزایش دهند. کودکان مبتلا به AA، به دلیل وجود نقایص ارثی متعدد و حساسیت به درمان‌ها، نیاز به تشخیص و رویکردهای درمانی متفاوت دارند.

تلومرها، که به حفاظت از انتهای کروموزوم‌ها کمک می‌کنند، در AA اهمیت دارند. طول تلومر می‌تواند به‌عنوان یک شاخص تشخیصی و درمانی استفاده شود. مطالعات ژنومی، تفاوت‌های قابل‌توجهی در بیان ژن‌های دخیل در متابولیسم سلول و سیستم ایمنی میان بیماران و افراد سالم نشان داده‌اند.

سلول‌های بنیادی مزانشیمی (MSCs) از مغز استخوان مشتق می‌شوند و نقش حیاتی در فرآیند خون‌سازی (هماتوپویز) و تنظیم سیستم ایمنی ایفا می‌کنند. این سلول‌ها به‌عنوان یک ریزمحیط حمایتی برای سلول‌های بنیادی خون‌ساز (HSCs) عمل می‌کنند و با ترشح فاکتورهای رشد و سیگنال‌دهی، بقای این سلول‌ها را تضمین می‌کنند. در بیماران مبتلا به کم‌خونی آپلاستیک (AA)، اختلالات عملکردی قابل‌توجهی در MSCs مشاهده شده است، از جمله کاهش ظرفیت تکثیر سلولی، کاهش توانایی در حمایت از HSCs، و تغییرات در بیان ژن‌های مرتبط با چرخه سلولی و سیستم ایمنی.

مطالعات نشان داده‌اند که MSCs در بیماران AA توانایی خود را در تنظیم فرآیندهای زیستی مانند تقسیم سلولی، تکثیر، و تمایز از دست داده‌اند. این اختلالات می‌تواند منجر به نقص در بازسازی مغز استخوان و کاهش تولید سلول‌های خونی شود. همچنین، تغییر در بیان ژن‌های دخیل در سیگنال‌دهی ایمنی، مانند ژن‌های مرتبط با چسبندگی سلولی و برقراری ارتباط سلولی، نشان می‌دهد که MSCs نمی‌توانند محیط مناسبی برای تنظیم سیستم ایمنی فراهم کنند. این نقص‌ها می‌تواند باعث افزایش پاسخ‌های التهابی و آسیب به سلول‌های مغز استخوان شود.

علاوه بر این، مطالعات بیان کرده‌اند که MSCs در بیماران AA دارای پتانسیل کاهش‌یافته در ترمیم آسیب‌های سلولی و مقابله با استرس اکسیداتیو هستند. این موضوع نشان می‌دهد که عملکرد ضعیف MSCs ممکن است یکی از عوامل کلیدی در ایجاد و پیشرفت کم‌خونی آپلاستیک باشد. با توجه به این یافته‌ها، تحقیقات بیشتری برای درک بهتر تغییرات عملکردی MSCs و نقش آن‌ها در بیماری لازم است. همچنین، MSCs به‌عنوان یک هدف بالقوه برای مداخلات درمانی، مانند پیوند سلول‌های بنیادی یا تنظیم ژنتیکی، مورد توجه قرار گرفته‌اند.

مطالعه حاضر به بررسی تاثیر مسیرهای نگهداری تلومر و وضعیت بیان ژن در MSCs می‌پردازد تا ارتباط آن‌ها با نتایج درمانی در کودکان مبتلا به AA تعیین شود.

۲- داده‌ها و روش پیش‌پردازش

این بخش توضیحی جامع در مورد روش‌های به‌کاررفته در مطالعه ارائه می‌دهد. تحقیق بر روی ۶ مجموعه‌داده که دارای بیماران مبتلا به کم‌خونی آپلاستیک و اهداکننده‌های سالم هستند، انجام شد. انتخاب بیماران مبتنی بر علائم کم‌خونی، کاهش سلول‌های خونی محیطی و مغز استخوان هیپوپلاستیک بوده است. افراد دارای سایر بیماری‌ها مانند HIV، هپاتیت، لنفوم، و بیماری‌های مزمن از مطالعه حذف شدند. نمونه‌های خون محیطی برای بررسی‌های مختلف جمع‌آوری شدند.

۲-۱- داده‌ها

مطالعه شامل سه گروه مختلف از بیماران بود: بیماران مبتلا به کم‌خونی آپلاستیک با طول تلومر طبیعی. بیماران مبتلا به کم‌خونی آپلاستیک با طول تلومر کاهش‌یافته. اهداکنندگان سالم.

DNA از نمونه‌های خون محیطی با استفاده از کیت‌های استاندارد استخراج شد. کیفیت و خلوص DNA با استفاده از اسپکتروفتومتر UV-Vis بررسی شد. برای اندازه‌گیری طول تلومر از واکنش زنجیره‌ای پلیمراز (qPCR) استفاده شد. برای این منظور، پرایمرهای خاصی برای تلومر و ژن‌های منفرد مورد استفاده قرار گرفتند. نتایج طبق روش استاندارد تفسیر شدند.

RNA کل از نمونه‌های خون محیطی استخراج شد و کیفیت آن با استفاده از ژل الکتروفورز تأیید شد. RNA استخراج‌شده برای سنتز cDNA با استفاده از ترانس‌کریپتاز معکوس مورد استفاده قرار گرفت.

بیان ژن‌های منتخب مرتبط با نگهداری تلومر با استفاده از روش qPCR ارزیابی شد. بیان ژن‌ها نسبت به GAPDH به‌عنوان کنترل داخلی محاسبه و با اهداکنندگان سالم کالیبره شد.

بیان ژن سلول‌های بنیادی مزانشیمی (MSCs) با استفاده از داده‌های ریزآرایه موجود در پایگاه داده Gene Expression Omnibus (GEO) انجام شد. این تحلیل شامل ۵ بیمار مبتلا به کم‌خونی آپلاستیک و ۵ اهداکننده سالم بود. تفاوت‌های بیان ژن با استفاده از تغییرات برابر یا بیشتر از دو برابر (upregulation) یا کمتر از نصف (downregulation) گزارش شدند.

دیتاست ۳۳۸۱۲GSE شامل پروفایل‌های بیان ژنی سلول‌های بنیادی مزانشیمی (MSCs) استخراج‌شده از مغز استخوان پنج بیمار کودکان مبتلا به کم‌خونی آپلاستیک شدید (SAA) است. این مطالعه با استفاده از میکروارای‌های الیگونوکلئوتیدی انجام شده و هدف آن بررسی بیان سیتوکین‌ها و فاکتورهای رشد ضروری در MSCها بوده است. نتایج نشان داد که پانزده ژن مرتبط با تکثیر سلولی و سیتوکین‌ها به‌طور قابل‌توجهی در بیماران SAA کاهش یافته‌اند. این یافته‌ها با استفاده از واکنش زنجیره‌ای پلیمراز معکوس کمی (qRT-PCR) تأیید شدند.

به‌طور خاص، کاهش بیان ژن‌های ۱۲CXCL (همچنین به‌نام ۱SDF) و فاکتور رشد هپاتوسیت (HGF) منجر به کاهش رشد سلولی و پاسخ به تمایز آدیپوژنیک، اما افزایش استئوژنز شد. بازگرداندن بیان ژن ۱۲CXCL یا افزودن HGF به‌صورت برون‌زا، عملکرد سلولی MSCهای بیماران SAA را تا حدی بازیابی کرد. این نتایج نشان می‌دهد که بیان ژن در MSCهای بیماران SAA به‌طور کلی کاهش یافته و ژن‌های ۱۲CXCL و HGF نقش‌های حیاتی در تنظیم رشد و تمایز سلولی ایفا می‌کنند.

دیتاست ۱۶۵۸۷۰GSE شامل داده‌های توالی‌یابی RNA (RNA-seq) از سلول‌های بنیادی و پیش‌ساز خونساز Lin-CD (HSPCs)۳۴+ استخراج‌شده از مغز استخوان سه اهداکننده سالم و بیماران مبتلا به کم‌خونی آپلاستیک (AA) درمان‌نشده است. برای تهیه این نمونه‌ها، ابتدا سلول‌های تک‌هسته‌ای از آسپیراسیون مغز استخوان با استفاده از جداسازی گرادیان چگالی فیکول جدا و در ۹۰٪ FBS و ۱۰٪ DMSO در نیتروژن مایع نگهداری شدند. پس از ذوب‌کردن، سلول‌ها با آنتی‌بادی‌های انسانی شامل کوکتل لاینئیج (۳CD، ۱۴CD، ۱۶CD، ۱۹CD، ۲۰CD، ۵۶CD) و ۳۴CD رنگ‌آمیزی شدند. سپس، سلول‌های HSPCs۳۴Lin-CD + با استفاده از دستگاه BD FACSAria III جداسازی و مستقیماً در بافر لیز سلولی قرار گرفتند. مراحل رونویسی معکوس و تعویض الگو بر اساس پروتکل ۲Smart-seq با برخی تغییرات انجام شد. توالی‌یابی با استفاده از پلتفرم Illumina NovaSeq ۶۰۰۰ و تولید خوانش‌های ۱۵۰ جفت‌باز انجام شد.

در پردازش داده‌ها، توالی‌های پرایمر TSO، دنباله پلی‌آ، آداپتور و توالی‌های با کیفیت پایین حذف شدند. توالی‌های تمیز به ژنوم انسانی ۳۸GRCh با استفاده از نرم‌افزار STAR هم‌تراز شدند و خوانش‌های تکراری با استفاده از Picard حذف شدند. برای تحلیل‌های بیشتر، از نرم‌افزارهای MATS و DaPars به ترتیب برای بررسی اتصالات جایگزین و پلی‌آدنیلاسیون جایگزین استفاده شد. کمی‌سازی ژن‌ها با استفاده از HTSeq انجام شد. این دیتاست به محققان امکان می‌دهد تا تفاوت‌های بیان ژنی بین HSPCs افراد سالم و بیماران AA را بررسی کنند و به درک بهتری از مکانیسم‌های مولکولی مرتبط با AA دست یابند.

دیتاست ۲۸۹۷۴GSE شامل پروفایل‌های بیان ژنی سلول‌های بنیادی مزانشیمی (MSCs) انسان است که با استفاده از پلتفرم Illumina Human-۶v ۲.۰ BeadChips مورد بررسی قرار گرفته‌اند. در این مطالعه، ۱۲ نمونه MSC مورد آنالیز قرار گرفته است.

این دیتاست بخشی از یک سوپر سری بزرگ‌تر است که شامل ۳۳۳۸۳GSE و ۴۲۳۵۲GSE می‌باشد و به شناسایی ژن‌های محرک استئوسارکوما از طریق تحلیل یکپارچه داده‌های تعداد نسخه و بیان ژن می‌پردازد. محققان می‌توانند از این داده‌ها برای درک بهتر مکانیسم‌های مولکولی مرتبط با MSCها و نقش آن‌ها در بیماری‌هایی مانند استئوسارکوما استفاده کنند.

دیتاست ۳۸۰۷GSE شامل پروفایل‌های بیان ژنی سلول‌های ۳CD + T استخراج‌شده از مغز استخوان و خون محیطی بیماران مبتلا به کم‌خونی آپلاستیک (AA) و داوطلبان سالم است. در این مطالعه، از ریزتراشه‌های Affymetrix HG_U۱۳۳A برای تحلیل بیان ژن‌ها استفاده شده است. نمونه‌ها شامل چهار بیمار AA (سه زن و یک مرد، ۱۹ تا ۷۰ ساله) بودند که از هر کدام نمونه‌های مغز استخوان و خون محیطی جمع‌آوری و RNA کل استخراج و ترکیب شد. همچنین، دو بیمار دیگر (یک مرد ۶۴ ساله و یک زن ۷۰ ساله) مورد بررسی قرار گرفتند که از آن‌ها نمونه‌های قبل و بعد از درمان جمع‌آوری شد. نمونه‌های کنترل سالم نیز از سه اهداکننده برای خون محیطی و دو اهداکننده برای مغز استخوان تهیه شدند.

هدف این مطالعه، مقایسه بیان ژنی در سلول‌های T بیماران AA و افراد سالم، و همچنین بررسی تغییرات بیان ژنی قبل و بعد از درمان در بیماران AA بود. این داده‌ها می‌توانند به درک بهتر مکانیسم‌های مولکولی مرتبط با AA و پاسخ به درمان کمک کنند. نتایج این تحقیق در مقاله‌ای با شناسه PubMed ۱۷۰۵۲۳۳۵ منتشر شده است.

دیتاست ۲۹۱۰۵GSE به بررسی پروفایل‌های بیان ژنی سلول‌های بنیادی خون‌ساز (HSCs) انسانی پرداخته است که از خون بند ناف استخراج شده‌اند. در این مطالعه، سلول‌ها بر اساس نشانگرهای سطحی ۳۴CD، ۳۸CD، RA۴۵CD، ۱Thy و f۴۹CD جداسازی شده‌اند تا جمعیت‌های مختلف HSCs و پیش‌سازهای چندتوان را شناسایی کنند. هدف اصلی این تحقیق، شناسایی ژن‌هایی است که در HSCها بیشتر بیان می‌شوند و ممکن است با عملکرد سلول‌های بنیادی و خودنوسازی آن‌ها مرتبط باشند.

برای انجام این آنالیز، RNA کل از جمعیت‌های سلولی جداسازی‌شده با استفاده از ریزتراشه‌های Illumina HumanHT-۱۲ V ۳.۰ expression

beadchip مورد بررسی قرار گرفته است. این داده‌ها می‌توانند به درک بهتر مکانیسم‌های مولکولی مرتبط با عملکرد HSCها کمک کنند و ژن‌های کلیدی مرتبط با خودنوسازی و تمایز این سلول‌ها را شناسایی نمایند. نتایج این مطالعه در مقاله‌ای با شناسه PubMed ۲۱۷۳۷۷۴۰ منتشر شده است.

دیتاست GSE113033 به بررسی پروفایل بیان ژنی زیرجمعیت‌های جدید سلول‌های بنیادی خون‌ساز (HSC) تعریف‌شده توسط بیان CD35 و جمعیت‌های پیش‌ساز در خون‌سازی بالغ انسان می‌پردازد. در این مطالعه، جمعیت‌های هدف با استفاده از دستگاه FACS Aria III جداسازی شدند و RNAاستخراج‌شده از این جمعیت‌ها با استفاده از میکروارای Agilent مورد آنالیز قرار گرفت. نتایج نشان داد که HSCهای +CD35 در مقایسه با HSCهای -CD35 و جمعیت‌های پیش‌ساز، بیان کمتری از ژن‌های مرتبط با چرخه سلولی یا تمایز دارند که نشان‌دهنده وضعیت خفته یا نابالغ آن‌ها است.

این دیتاست شامل ۱۸ نمونه است که در پلتفرم Agilent-۰۷۲۳۶۳ SurePrint G ۳Human GE v ۳ ۸x۶۰ Microarray ۰۳۹۴۹۴K مورد بررسی قرار گرفته‌اند. نمونه‌ها شامل تکرارهای بیولوژیکی از HSCهای CD۳۵+، HSCهای CD۳۵-، سلول‌های پیش‌ساز چندتوانی (MPP)، پیش‌سازهای میلوئیدی مشترک (CMP)، پیش‌سازهای گرانولوسیت–ماکروفاژ (GMP) و پیش‌سازهای مگاکاریوسیت–اریتروایِد (MEP) هستند. این مطالعه توسط محققان دانشگاه کیوشو در ژاپن انجام شده است.

۲-۲- روش پیش‌پردازش

با توجه به استفاده از مجموعه‌داده های متفاوت که هر کدام تکنولوژی مختلفی دارند و همچنین پردازش داده های مجموعه داده مقاله به شکل خام[۲] روش‌های پیش پردازش گوناگونی را انجام دادیم که به شرح زیر هستند:

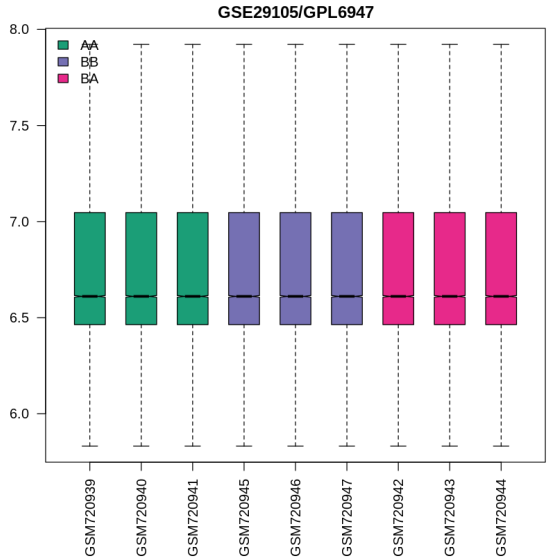
۲-۲-۱- داده‌های موجود در مقاله اصلی

خانم صادق آبادی

۲-۲-۲- داده‌های افزوده مشابه مقاله اصلی به شکل مستقل از هم

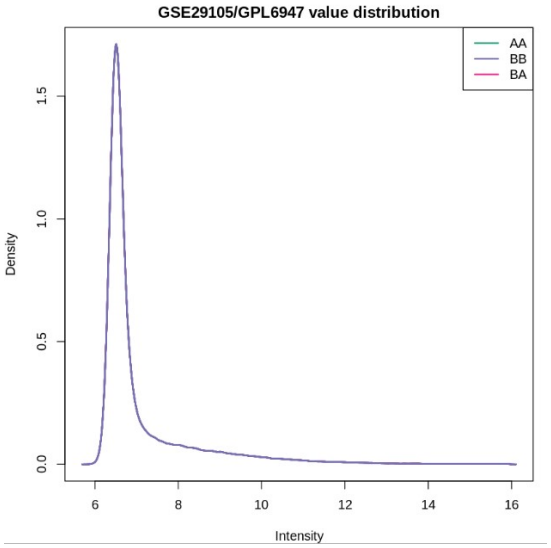
کد های کامل پیش‌پردازش و تحلیل این مجموعه داده ها به صورت مجزا در نوتبوک های ژوپیتر ضمیمه شده و با کامنت های شفاف کننده آمده اند. آنها را می توان در فایل های GSE28974.ipynb، GSE3807.ipynb، GSE29105.ipynb، GSE113033.ipynb و GSE165870.ipynb مشاهده کرد. در اینجا به طور مختصر آنها را توضیح داده و سپس خروجی های موجود در یکی از آنها را میبینیم. به طور خاص مجموعه داده GSE29105 را در اینجا بررسی میکنیم و توجه داریم که ۴ مجموعه داده دیگر نیز مشابه همین مجموعه داده هستند.

- در کد R موجود مشاهده می شود که ابتدا به نصب package های ضروری نظیر DESeq2، GEOquery و غیره می پردازیم. سپس کتابخانه های ضروری را import میکنیم.
- در مرحله بعد داده ها را از پایگاه داده GEO دانلود میکنیم. سپس نام ستون ها را با استفاده از فایل annotation اصلاح میکنیم.
- حال به گروه بندی نمونه ها با توجه به توضیحات خاص مجموعه داده میپردازیم.
- حال که ماتریس بیان ژن ها در نمونه ها را داریم، به نرمال سازی داده ها با استفاده از روش quantile normalization می پردازیم. سپس log2 transformation را انجام می دهیم. در نهایت به مدیریت missing values و حذف یا پر کردن آنها با استفاده از روش مناسب نظیر مقدار میانه می پردازیم.
- در مرحله بعد نمودار جعبه ای را رسم میکنیم تا از نرمال سازی درست مطمئن شویم که در شکل زیر مشاهده میکنید. همان طور که میبینید داده ها به خوبی نرمال شده اند.
- در نهایت نیز، با رسم نمودار توزیع مقادیر و نمونه های کاهش ابعاد یافته نمونه ها با الگوریتم unmap از درستی کد خود اطمینان حاصل میکنیم.

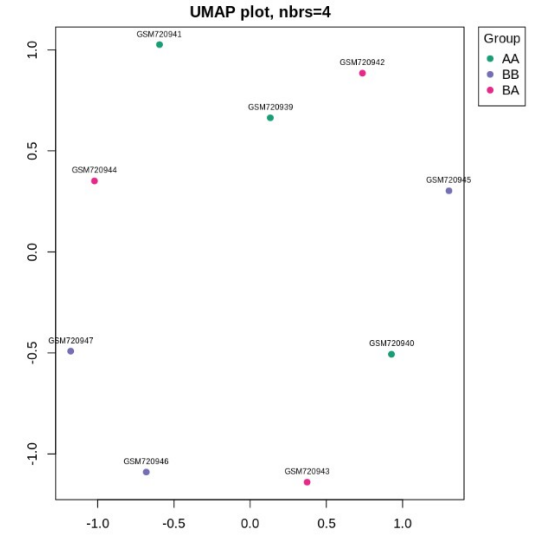


شکل (۱) : نمودار جعبه ای پس از نرمال سازی

[۲] raw



شکل (۲) : نمودار توزیع مقادیر پس از پیش پردازش



شکل (۳) : نمودار کاهش ابعاد نمونه ها بعد از پیش پردازش

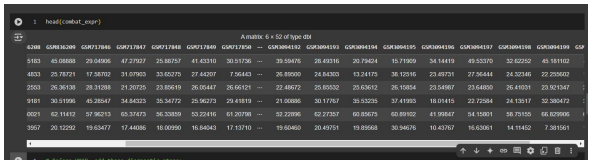
۳-۲-۲- داده‌های افزوده مشابه مقاله اصلی به شکل ترکیب شده[۳]

کد های کامل پیش‌پردازش و تحلیل این مجموعه داده ها به صورت ترکیب شده و یکجا در نوتبوک ژوپیتر ضمیمه شده و با کامنت های شفاف کننده آمده است. آن را می توان در فایل bio_proj_merged_5_datasets.ipynb مشاهده کرد. توجه داریم که ۵ مجموعه داده را با هم ترکیب کردیم که یکی از آنها مجموعه داده اصلی مقاله خواسته شده است. ۴ مجموعه داده دیگر مجموعه داده های مشابه آن هستند. همچنین برای این قسمت چون تکنولوژی مجموعه داده GSE165870 از نوع RNA-seq data است از آن صرف نظر کردیم چرا که سایر مجموعه داده ها دارای تکنولوژی Microarray هستند. در اینجا به طور مختصر آن را توضیح داده و سپس خروجی های موجود در آن را میبینیم.

[۳] merged

توجه داریم چالش اصلی در این قسمت این است که داده های هر مجموعه داده دارای فناوری و platform های گوناگونی هستند. این منجر به تفاوت نامگذاری سطر های ماتریس بیان ژن برای نام ژن ها و همچنین ژن های غیر مشترک موجود می شود.

- در کد R موجود مشاهده می شود که ابتدا به نصب package های ضروری نظیر limma، GEOquery و غیره می پردازیم. سپس کتابخانه های ضروری را import میکنیم.
- با توجه به چالش فناوری های مختلف هر مجموعه داده یک تابع با نام process_dataset تعریف کردیم که بعد از دانلود هر مجموعه داده با توجه به فناوری مجموعه داده و annotation های موجود همه مجموعه داده ها را به یک فضای نامگذاری نگاشت کند.این تابع پیچیده ای است که توضیح جزییات آن در اینجا نمی گنجد ولی با مراجعه به نوتبوک ضمیمه شده و خواندن کامت ها می توانید جزییات آن را مشاهده کنید.
- سپس، با تابع process_dataset این ۵ مجموعه داده را دانلود و پردازش میکنیم. همچنین پس از اینکه نام ژن ها یکی شد با اشتراک گرفتن بین ۵ مجموعه داده، ژن های مشترک آن ها را پیدا میکنیم.
- در مرحله بعد، یک ماتریس که حاصل ترکیب این ۵ مجموعه داده است می سازیم که سطر های آن ژن های مشترک و ستون ها آن نمونه های مختلف در مجموعه داده ها هستند.
- حال با کتابخانه ComBat عملیات Batch correction را انجام می دهیم که بسیار ضروری است چرا که مجموعه داده ها فناوری های گوناگونی دارند.
- در نهایت نیز مقادیر NA/NaN را از ماتریس خود حذف میکنیم. در شکل زیر میتوانید ۶ سطر اول ماتریس حاصل از ترکیب مجموعه داده ها مشاهده کنید.



شکل (٤) : مجموعه داده ها پس از ترکیب و پیش پردازش

۳- نتایج

همون طور که در بخش قبل توضیح دادیم ما به سه شیوه مختلف مجموعه داده ها را ارزیابی کردیم. لذا نتایج هر قسمت و تحلیل های آنها را در ادامه و به شکل تفکیک شده مشاهده میکنید.

۱-۱-۳- داده‌های موجود در مقاله اصلی

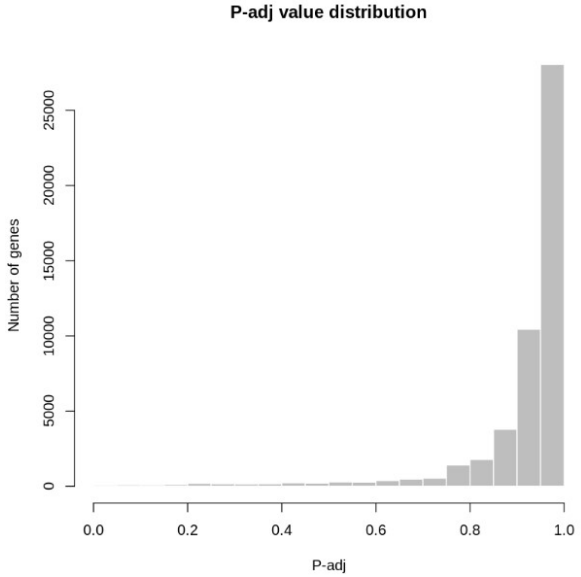
خانم صادق آبادی

۲-۱-۳- داده‌های افزوده مشابه مقاله اصلی به شکل مستقل از هم

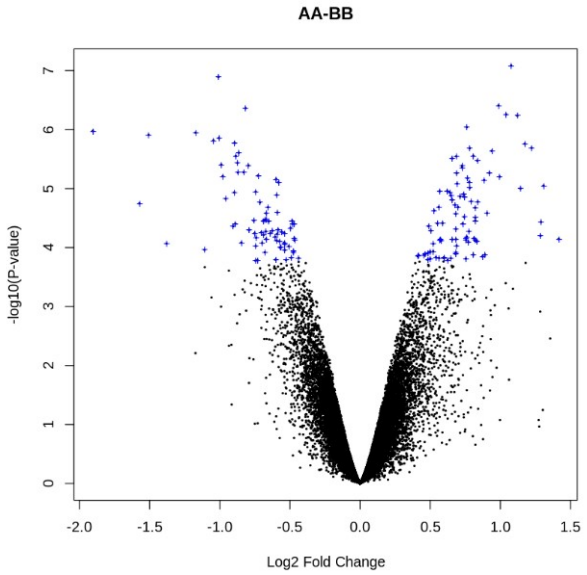
نمودار و خروجی های به دست آمده به شکل کامل در فایل های GSE29105.ipynb، GSE3807.ipynb، GSE28974.ipynb، GSE113033.ipynb و GSE165870.ipynb قابل مشاهده هستند. در

اینجا به توضیح مختصری از یکی از آنها می پردازیم چرا که شبیه به هم بوده و همچنین در صورت توضیح هر ۵ مجموعه داده به شکل جدا جـدا گـزارش طولانی خواهد شد و انسجام خـود را از دسـت مـی دهـد. . بـه طـور خـاص مجموعه داده GSE29105 را در اینجا بررسی میکنیم و توجه داریـم کـه ۴ مجموعه داده دیگر نیز مشابه همین مجموعه داده هستند.

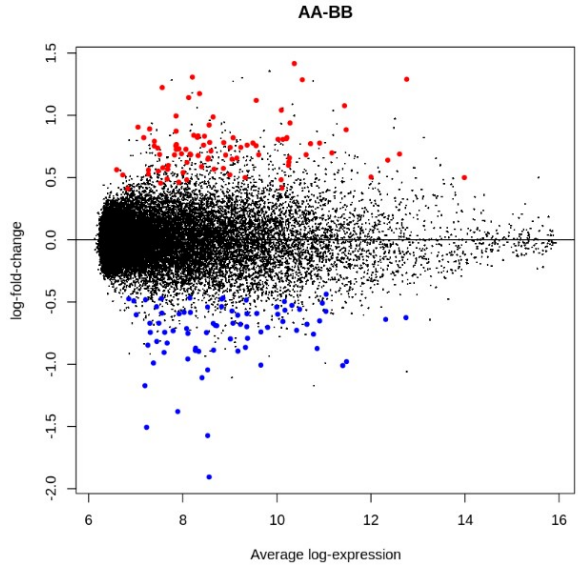
برای محاسبه نتایج ابتدا یک مدل خطی را روی مجموعه داده پیش پـردازش شده fit میکنیم. سپس، با استفاده از مدل بیزی و آزمون های آماری، آماره ها و مقـادیر لازم بـرای تحلیـل را بـه دسـت مـی‌آوریم از قبیـل P.Value، adj.P.Val، logFC و غیره. در نهایت نیز با استفاده از مقادیر به دست آمده و همچنین کشیدن نمودار های مختلف ژن هایی که در گـروه هـای مختلـف اهمیت دارند و differentially بین شده اند را شناسایی میکنیم. در ادامه در شکل های زیر نمودار توزیع مقادیر adj.P.Val، نمودار آتش فشانی و نمودار MD را مشاهده میکنید.



شکل (۵) : نمودار توزیع مقادیر adj.P.Val برای مجموعه داده GSE29105



شکل (٦) : نمودار آتش فشانی گروه AA در برابر BB برای مجموعه داده GSE29105



شکل (۷) : نمودار MD گروه AA در برابر BB برای مجموعه داده GSE29105

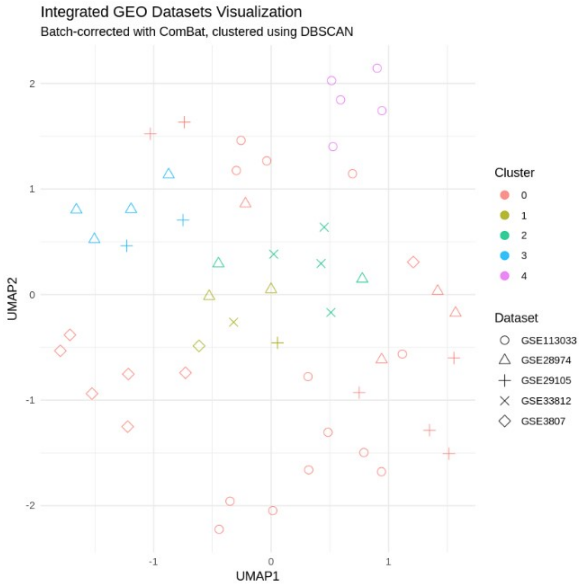
۳-۱-۳- داده‌های افزوده مشابه مقاله اصلی به شکل ترکیب شده

نمودار و خروجی های به دست آمده به همراه کد های لازم به شکل کامل در فایل bio_proj_merged_5_datasets.ipynb قابل مشاهده اسـت. در اینجا به توضیح مختصری از آن می پردازیم.

برای ارزیابی شباهت و اطمینان از کیفیت مجموعـه داده هـای پیـدا شـده بـا مجموعه داده اصلی موجود در مقاله و meta-analysis این بخش برای مـا اهمیت دارد.

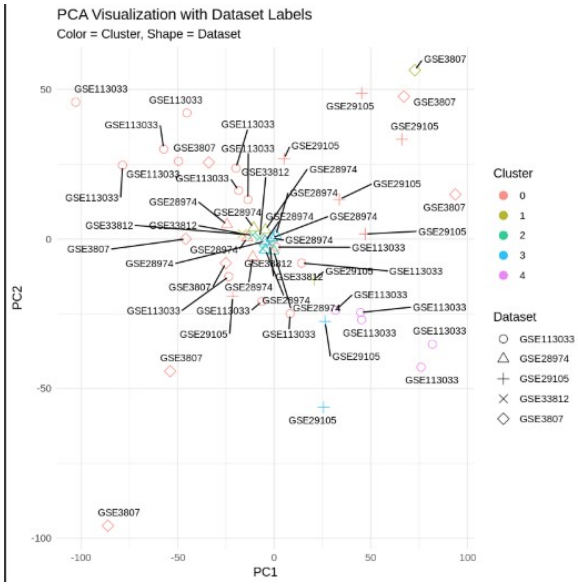
ابتدا از scale مناسب داده های اطمینان حاصل میکنیم چرا که الگوریتم های بعدی از قبیل کاهش ابعاد به داده های نرمال شده نیاز دارند. سپس ژن هـای با واریانس صفر را حذف میکنیم چرا که هیچ اطلاعات مفیدی برای ما ندارند. حال از الگوریم کاهش ابعـاد UMAP بـا ابرپـارامتر هـای مناسـب اسـتفاده میکنیم. سپس برای مصورسازی بهتر از یـک الگـوریتم خوشـه بنـدی ماننـد DBSCAN استفاده میکنیم.

نتیجه به دست آمده را در شکل زیر مشاهد میکنید. همـان طـور کـه میبینیـد نمونه های موجود در هر مجموعه داده به هم نزدیـک تـر هسـتند. همچنـین مجموعه داده اصلی و مرجع ما یعنی GSE33812 در مرکز قرار گرفتـه کـه نشان دهنده نزدیکی نمونه های آن به سایر مجموعه داده ها است که گواه بر کیفیت و نزدیکی سایر مجموعه داده ها به آن هسـت. همچنـین مـی تـوانیم مجموعه داده های همسایه را ببینید که نشان می دهد کدام مجموعه داده ها قرابت بیشتری به هم دیگر دارند. نتیجـه جالـب و مـورد انتظـار دیگـر اینکـه مجموعه داده GSE113033 از بقیه فاصله بیشـتری دارد کـه بـا توجـه بـه ماهیت زیستی آن و توضیحات مجموعه داده کاملا مورد انتظار است.

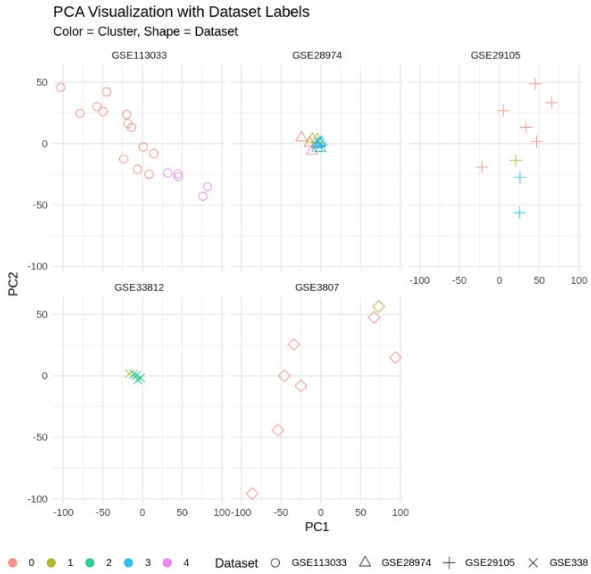


شکل (۸) : نمودار کاهش ابعاد با الگوریتم UMAP و اجرای الگوریتم خوشه بندی رو مجموعه داده های ترکیب شده با هم

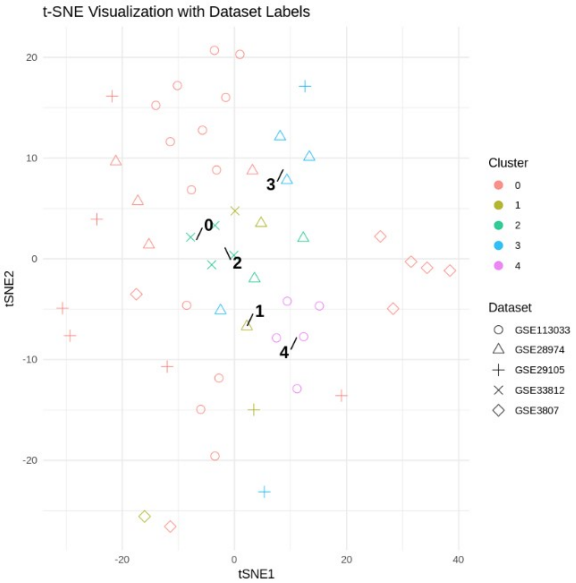
همچنین برای تحلیل جامع تر از الگوریتم های کاهش ابعاد PCA و t-SNE نیز استفاده کردیم و نمودار های زیر را کشیدیم که تایید کننـده توضـیحات و نتایج بالا هستند.



شکل (۹) : نمودار کاهش ابعاد با الگوریتم PCA و اجرای الگوریتم خوشه بندی رو مجموعه داده های ترکیب شده با هم



شکل (۱۰) : نمودار کاهش ابعـاد بـا الگـوریتم PCA و اجـرای الگوریتم خوشه بندی رو مجموعه داده های ترکیب شده با هـم به تفکیک مجموعه داده



شکل (۱۱) : نمودار کاهش ابعاد با الگوریتم t-SNE و اجرای الگوریتم خوشه بندی رو مجموعه داده های ترکیب شده با هم

۴- نتیجه گیری

کودکان مبتلا به آنمی آپلاستیک در صورت دریافت درمان مناسب، شامل درمان‌های سرکوب‌کننده ایمنی و پیوند مغز استخوان، نتایج درمانی امیدوارکننده‌ای خواهند داشت. طول تلومر لکوسیت‌ها و وضعیت بیان ژنی ژن‌های GAS۲TERF, MKI۳L۲, ۶۷ و A۱۵TMSB به عنوان یک نشانگر زیستی بالقوه برای گروه‌های مختلف از بیماران کودکان مبتلا به آنمی آپلاستیک شناسایی شده است. بر اساس این آزمایش‌های نشانگر زیستی، گزینه‌های درمانی مناسب برای این اختلال قابل انتخاب خواهند بود.

مراجع

[1] S.A. Peslak, T. Timothy Olson, D.V. Babushok, Diagnosis and treatment of aplastic anemia, Curr. Treat. Options Oncol. 18 (12) (2017) 70.

[2] N.S. Young, D.W. Kaufman, The epidemiology of acquired aplastic anemia, Haematologica 93 (4) (2008) 489–492.

[3] C. Dufour, P. Veys, E. Carraro, N. Bhatnagar, M. Pillon, R. Wynn, et al., Similar outcome of upfront-unrelated and matched sibling stem cell transplantation in idiopathic paediatric aplastic anaemia, Br. J. Haematol. 171 (4) (2015) 585–594.

[4] Kaufman DW, Kelly JP, Levy M, Shapiro S. The Drug Etiology of Agranulocytosis and Aplastic Anemia 1991; Oxford University Press, New York.

[5] L.E. Böttiger, B. Westerholm, Aplastic anaemia: I. Incidence and aetiology, Acta Med. Scand. 192 (1972) 315–318.

[6] S.M. Davies, D.J. Walker, Aplastic anaemia in the Northern Region 1971–1978 and follow-up of long term survivors, Clin. Lab. Haematol. 8 (1986) 307–313.

[7] M. Szklo, L. Sensenbrenner, J. Markowitz, S. Weida, S. Warm, M. Linet, Incidence of aplastic anemia in metropolitan Baltimore: a population-based study, Blood 66 (1985) 115–119.

[8] C. Isaragrisl, C. Sriratanasatavorn, A. Piankijagum, S. Vannasaeng, Y. Porapaklham, P.E. Leaverton, et al., Incidence of aplastic anemia in Bangkok, Blood 77 (1991) 2166–2168.

[9] N.S. Young, Aplastic anemia, Lancet 346 (1995) 228–232.

[10] N.S. Young, R. Calado, P. Scheinberg, Current concepts in the pathophysiology and treatment of aplastic anemia, Blood 108 (2006) 2509–2519.

[11] R.A. Brodski, R.J. Jones, Aplastic anaemia, Lancet 365 (2005) 1647–1656.

[12] N. Yoshida, H. Yagasald, A. Hama, Y. Takahashi, Y. Kosaka, R. Kobayashi, et al., Predicting response to immunosuppressive therapy in childhood aplastic anemia, Haematologica 96 (2011) 771–774.

[13] F. Timeus, N. Crescenzio, A. Lorenzati, et al., Paroxysmal nocturnal hemoglobinuria clones in children with acquired aplastic anemia: a prospective single centre study, Br. J. Haematol. 150 (2010) 483–485.

[14] L.C. Fox, D.S. Ritchie, Pediatric aplastic anemia treatment patterns and responses; power in the numbers, Haematologica 104 (10) (2019) 1909–1912.

[15] E.H. Blackburn, Switching and signaling at the telomere, Cell 106 (6) (2001) 661–673.

[16] H. Sakaguchi, N. Nishio, A. Hama, N. Kawashima, X. Wang, A. Narita, S. Doisaki, Y. Xu, H. Muramatsu, N. Yoshida, Y. Takahashi, K. Kudo, H. Moritake, K. Nakamura, R. Kobayashi, E. Ito, H. Yabe, S. Ohga, A. Ohara, S. Kojima, Japan Childhood Aplastic Anemia Study Group, Peripheral blood lymphocyte telomere length as a predictor of response to immunosuppressive therapy in childhood aplastic anemia, Haematologica 99 (8) (2014) 1312–1316.

[17] B. Hubner, S. Merk, S. Rauhut, M. Dugas, T. Haferlach, M. Fuehrer, A. Borkhardt, Individual gene expression profiling of bone marrow CD34 cells in acquired severe aplastic anemia (aSAA) in children, Blood 108 (11) (2006) 978.

[18] J. Li, S. Yang, S. Lu, H. Zhao, J. Feng, W. Li, F. Ma, Q. Ren, B. Liu, L. Zhang, et al., Differential gene expression profile associated with the abnormality of bone marrow mesenchymal stem cells in aplastic anemia, PLoS One 7 (2012) e47764.

[19] M.C. Kastrinaki, K. Pavlaki, A.K. Batsali, E. Kouvidi, I. Mavroudi, C. Pontikoglou, H.A. Papadaki, Mesenchymal stem cells in immune-mediated bone marrow failure syndromes, Clin. Dev. Immunol. 2013 (2013) 265608.

[20] Y.H. Chao, C.T. Peng, H.J. Ham, C.K. Chan, K.H. Wu, Poor potential of proliferation and differentiation in bone marrow mesenchymal stem cells derived from children with severe aplastic anemia, Ann. Hematol. 89 (2010) 715–723.

[21] E. Hamzic, K. Whiting, E. Gordon Smith, R. Pettengell, Characterization of bone marrow mesenchymal stromal cells in aplastic anaemia, Br. J. Haematol. 169 (2015) 804–813.

[22] S. Fujimaki, H. Harigae, T. Sugawara, et al., Decreased expression of transcription factor GATA-2 in haematopoietic stem cells in patients with aplastic anaemia, Br. J. Haematol. 113 (2001) 52–57.

[23] Y.H. Chao, K.H. Wu, S.H. Chiou, et al., Downregulated CXCL12 expression in mesenchymal stem cells associated with severe aplastic anemia in children, Ann. Hematol. 94 (2015) 13–22.

[24] Y. Xu, Y. Takahashi, Y. Wang, et al., Downregulation of GATA-2 and overexpression of adipogenic gene PPAR-

[39] S. Adbikari, P. Mandal, Integrated analysis of global gene and microRNA expression profiling associated with aplastic anaemia, Life Sci. 228 (2019) 47–52.

gamma in mesenchymal stem cells from patients with aplastic anemia, Exp. Hematol. 37 (2009) 1393–1399.

[25] E.P. Weinzierl, D.A. Arber, The differential diagnosis and bone marrow evaluation of new-onset pancytopenia, Am. J. Clin. Pathol. 139 (1) (2013) 9–29.

[26] R.M. Cawthon, Telomere measurement by quantitative PCR, Nucleic Acids Res. 30 (10) (2002) e47.

[27] N. O'Callaghan, V. Dhillon, P. Thomas, M. Fenech, A quantitative real-time PCR method for absolute telomere length, Biotechniques 44 (6) (2008) 807–809.

[28] B. Sebastian, Protocol GMBOO3:B: 20090216NP, Genomic Medicine Biorepository GMB003, Revised and Approved 2012JULY26 by B. Sebastian.

[29] K. Chow, et al., Gene Expression Profiles of Bone Marrow Mesenchymal Stem Cells in Pediatric Patients with Severe Aplastic Anemia, 2011 (Chow K. et al., 2011, accession GSE33812).

[30] F. Beier, M. Foronda, P. Martinez, M.A. Blasco, Conditional TRF1 knockout in the hematopoietic compartment leads to bone marrow failure and recapitulates clinical features of dyskeratosis congenita, Blood 120 (15) (2012) 2990–3000.

[31] J.L. Cheng, A.L. Wang, J. Wan, Association between the M235T polymorphism of the AGT gene and cytokines in patients with hypertension, Exp. Ther. Med. 3 (3) (2011) 509–512.

[32] K. Matsushita, Y. Wu, Y. Okamoto, R.E. Pratt, V.J. Dzau, Local renin angiotensin expression regulates human mesenchymal stem cell differentiation to adipocytes, Hypertension (Dallas, Tex.: 1979) 48 (6) (2006) 1095–1102.

[33] W.X. Carroll, N.S. Kalupahana, S.L. Booker, N. Siriwardhana, M. Lemieux, A.M. Saxton, et al., Angiotensinogen gene silencing reduces markers of lipid accumulation and inflammation in cultured adipocytes, Front. Endocrinol. 4 (2013) 10.

[34] S. Stopp, M. Grindl, M. Fackler, J. Malkmus, M. Leone, R. Naumann, S. Frantz, E. Wolf, B. von Eyss, F.B. Engel, S. Gaubatz, Deletion of Gas2l3 in mice leads to specific defects in cardiomyocyte cytokinesis during development, Proc. Natl. Acad. Sci. U.S.A. 114 (2017) 8029–8034.

[35] S. Kordasti, B. Costantini, T. Seidl, P.P. Abellan, M.M. Llordella, D. McLornan, K. Diggins, A. Kulasekararaj, C. Benfatto, X. Feng, A. Smith, S.A. Mian, R. Melchiotti, E. de Rinaldis, R. Ellis, N. Petrov, G.A.M. Powles, S.S. Chung, N.S. Thomas, F. Farzaneh, J.M. Irish, S. Heck, N.S. Young, J.C.W. Marsh, G.J. Mufti, Deep phenotyping of Tregs identifies an immune signature for idiopathic aplastic anemia and predicts response to treatment, Blood 128 (9) (2016) 1193–1206.

[36] P. Bertheau, E. Turpin, D.S. Rickman, M. Espie, A. de Reynies, J.P. Feugeas, L. Plassa, H. Soliman, M. Varna, A. de Roquancourt, J. Lehmann-Che, Y. Beuzard, M. Marty, J.L. Misset, A. Janin, H. de The, Exquisite sensitivity of TP53 mutant and basal breast cancers to a dose-dense epirubicin-cyclophosphamide regimen, PLoS Med. 4 (2007) e90.

[37] C.N.N. Weiss, K. Ito, A macro view of MicroRNAs: the discovery of MicroRNAs and their role in hematopoiesis and hematologic disease, Int. Rev. Cell Mol. Biol. 334 (2017) 99–175.

[38] E.C. Guinan, Aplastic anemia: management of pediatric patients, Hematology Am. Soc. Hematol. Educ. Program (2005) 104–109.