استفاده از توالی‌یابی long-read برای روشن کردن پیچیدگی‌های فارماکوژنومیک

گلناز زمانیان [1]

۱- کارشناسی ارشد سم شناسی ، دانشگاه علوم پزشکی تهران ، تهران، ایران
پژوهشگر مرکز تحقیقات پزشکی شخصی آمیتیس

مقدمه

فارماکوژنومیک (PGx) برای فردی کردن دوزهای دارو و در نتیجه بهبود نتایج درمان دارویی بسیار مهم است. PGx به فنوتیپ‌های استنباط شده بر اساس واریانت‌های شناخته شده در فارماکوژن‌ها متکی است. با این وجود، واریانت ژنتیکی در پاسخ به دارو و فعالیت آنزیم را نمی توان با سنجش‌های ژنتیکی معمول PGx، به دلیل عوامل متعدد، توضیح داد. اول، سنجش‌های ژنوتیپ‌سازی کنونی قادر به شناسایی کامل ساختار ژنتیکی همه ژن‌های دخیل در پاسخ دارویی نیستند. دوم، مکانیسم اثر دارو و یا مسیر متابولیک آن همیشه به طور کامل شناخته نشده است. برای ارزیابی اینکه چه بخشی تنوع ژنتیکی است و چه بخشی می‌تواند توسط عوامل دیگر توضیح داده شود، ضروری است که بتوانیم تمام مولفه‌های ژنتیکی را که باعث پاسخ دارویی متغیر می‌شوند توضیح دهیم. با این حال، این مورد دارای چالش است زیرا اکثر فارماکوژن‌ها حداقل تا حدی در مناطق پیچیده ژنومی قرار دارند یا دارای واریانت‌های مختلفی مانند تکرارهای پشت سر هم و ترکیبات هیبریدی شبه ژن هستند. فن‌آوری‌های ژنوتیپ‌سازی در حال حاضر بر اساس ریزآرایه‌های SNV (Single Nucleotide Variant) یا توالی‌یابی کوتاه خوانده می‌شوند. هر دو رویکرد در توصیف این مناطق پیچیده محدود هستند، زیرا آنها قادر به حل و فصل مناسب و قابل اعتماد مناطق بسیار همولوگ و شناسایی واریانت PGx نیستند. علاوه بر این، با فازبندی هاپلوتیپ می‌توان تعیین کرد که آیا واریته‌ها روی یک آلل قرار دارند یا روی آلل‌های متفاوتی قرار دارند، که به طور بالقوه منجر به تفاوت در انتساب فنوتیپ می‌شود. در حال حاضر هاپلوتیپ‌های PGx بر اساس عدم تعادل

پیونـــدی مرحله‌بندی می‌شـــوند. در حالی که این منجر به هاپلوتیپ‌های دقیق در مقیاس جمعیتی می‌شـــود، اما همیشه به فرضیات دقیق در سطح فردی منجر نمی‌شود. تأثیر این چالش‌ها در عمل بالینی زیاد اســـت. به عنوان مثـــال، ژن پیچیده CYP2D6 در متابولیســـم ۲۰ تا ۳۰ درصد داروهای معمول تجویز می‌شـــود و نمی‌توان آن را به طور کامل با توالی‌یابی کوتاه مشخص کرد.

در ســـال‌های اخیـــر، فناوری‌های توالی‌یابـــی طولانی از Oxford Nanopore و PacBio نشان داده‌اند که می‌توانند مناطق پیچیده ژنومی (دارویی) را مشـــخص کنند. برای این مناطق، خوانش طولانی و با کیفیت بالا به طور قابل توجهی دقت جســـتجو انواع را بهبود می‌بخشد و امکان تفکیک دیپلوتیپ‌های کاملاً فازی را فراهم می‌کند.

ارزش توالی‌یابـــی طولانی مدت برای اهداف تشـــخیصی بیماری قبلاً نشان داده شده اســـت. توالی‌یابی PacBio نشـــان داده شـــده اســـت که می‌توانـــد CYP2D6 را با پوشـــش کل مکان ژن در یک خوانش طولانی با کیفیت بالا مشـــخص کند. اخیراً، توالی‌یابـــی طولانی نیز برای ژن‌های HLA در رابطه با PGx اعمال شـــده است. علاوه بر این، کاربرد آن در بســـیاری از سنجش‌های تحقیقاتی تشخیصی بالینی چالش برانگیز مانند تکرارهای طولانی مدت در ژن FMR1 مرتبط با سندرم X شکننده و در ژن PKD1 برای شناسایی جهش‌های مرتبط با بیماری کلیه پلی کیســـتیک استفاده شده است. در نهایت، توالی‌یابی، مرحله‌بنـــدی هاپلوتیپ را بـــدون نیاز بـــه رویکردهای محاسباتی و یا اطلاعات شجره نامه تسهیل می‌کند. این می‌تواند در PGx اهمیت بســـیار زیادی داشته باشد که منجر به پیش بینی‌های دقیق‌تر فنوتیپ می‌شود. ترکیب پیچیدگی PGx و فازبندی هاپلوتیپ نشـــان می‌دهد که توالی‌یابی طولانی پتانســـیل این را دارد که به طور قابل توجهـــی توانایی ما در پیش بینـــی صحیح فنوتیپ‌های متابولیزه کننده دارو را بهبود بخشد. در این مقاله اثبات مفهوم، ما پتانســـیل توالی‌یابی طولانی‌مدت PacBio را برای حـــل نواحی پیچیده PGx با اســـتفاده از داده‌های توالی‌یابی موجود ژنوم به خوبی مشخص‌شـــده در نمونه مرجع HG002 بطری (GIAB) ارزیابی می‌کنیم.

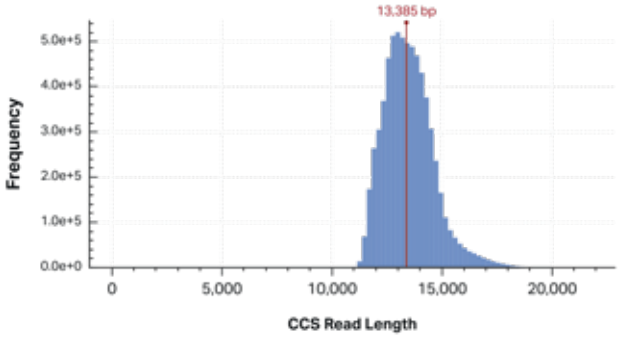
نتایج

تعریف داده ها

داده‌های توالی‌یابی قبلی منتشـــر شده از نمونه HG۰۰۲ GIAB به خوبی مشخص شـــده به دست آمد. این داده شـــامل ۶,۷۲۸,۱۲۳ خواهش با طول متوسط ۱۳.۴ اینچ کیلوبایت است که ۹۷.۵ درصد از ژنوم با میانگین پوشش نقشـــه برداری شده ۲۸ برابری را پوشش می‌دهد (شکل ۱). تقریباً ۵ میلیون واریانت ژنتیکی با استفاده از GATK HaplotypeCaller و DeepVariant شناسایی شد.

دقت و جستجو بالا در جستجو واریانت ها

برای ۱۰۰ فارماکوژن انتخاب شـــده، دقت و یادآوری در مقایســـه با مجموعه حقیقت معیار GIAB v3.3 تعیین شـــد. بـــرای SNVهـــا، GATK HaplotypeCaller و DeepVariant به دقت و جســـتجو مشابهی بالای ۹۹.۸ درصد دســـت یافتند (جـــدول ۱). با ایـــن حال، تماس گیرنده DeepVariant عملکرد بسیار بهتری در تشخیص ایندل (۹۸٪ <) در مقایســـه با GATK (دقت: ۹۴.۵٪ و یادآوری: ۸۶.۱٪) به دســـت آورد. هنگام مقایسه با نتایج گسترده ژنوم گزارش شده توسط ونگر و همکاران، دقت و یادآوری در تشـــخیص واریانت‌هـــای فارماکوژن‌ها برتر اســـت. هنگام طبقه بندی نتایج در مناطق پیچیده، دقت بالا باقی ماند، با یادآوری و دقت بیش از ۹۵٪ برای همه مناطـــق برای هر دو indels و SNV. برای تماس گیرنده GATK، دقت کمتر بود (۱۰۰٪-۸۵ در مقایســـه با ۹۷-۱۰۰٪ برای تماس گیرنده DeepVariant). کاهش دقت



شکل ۱: توزیع طول خوانش
توزیع طول خوانده شده ژنوم در نمونه بطری HG002 پس از توالی‌یابی بر روی پلت فرم دنباله علوم زیستی اقیانوس آرام و ساخت توالی اجماع دایره‌ای.

را می‌توان به عملکرد پایین‌تر برای تکرارهای پشت سر هم و هموپلیمرها نسبت داد.

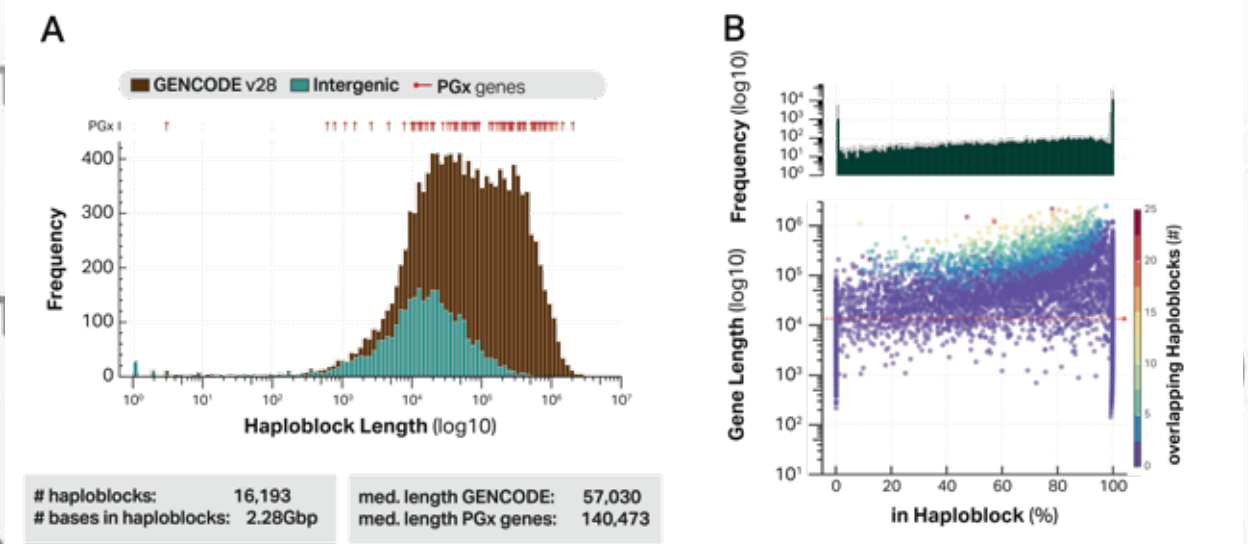
برای ارزیابی دقت جستجو SV در فارماکوژن‌ها، جستجوهای SV با معیار SV تعیین شده برای همه SVهای بیش از ۵۰ جفت باز مقایسه شد. با این حال، مناطق GIAB با اطمینان بالا همه ۱۰۰ ژن را پوشش نمی‌دهند. ۴۶ ژن حذف شدند، ۱۲ ژن تا حدی و ۴۲ ژن به طور کامل با داده‌های انتخاب شده GIAB همپوشانی داشتند. در مجموع، ۲۲ SV (بیش از ۵۰ جفت باز) در ۵۴ فارماکوژن در مقایسه با ۲۳ فهرست شده در مجموعه معیار شناسایی شد. دو تماس منفی کاذب و یک تماس مثبت کاذب در نظر گرفته شد. با هم، ارزیابی عملکرد تشخیص SVs در مناطق PGx منجر به جستجو ٪۹۱.۳ و دقت ٪۹۴.۵ شد. یادآوری و دقت بالا در فارماکوژن‌ها نشان می‌دهد که با استفاده از داده‌های توالی‌یابی طولانی در مقایسه با معیارهای فعلی، دقت از دست نمی رود، در حالی که تشخیص انواع پیچیده ژنتیکی را بهبود می‌بخشد.

فازبندی هاپلوتیپ و هاپلوبلاک ها

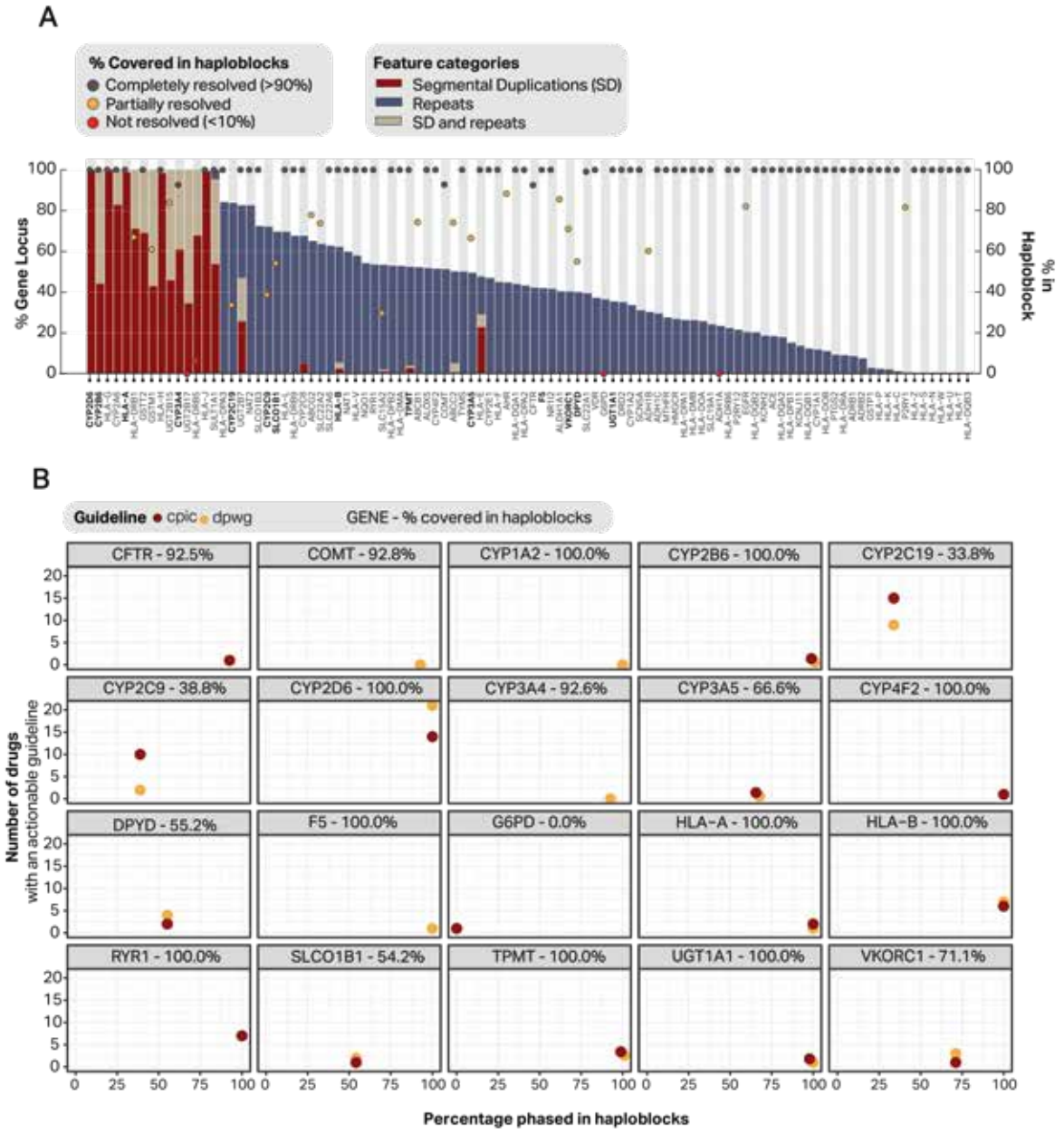
با استفاده از واتس‌هپ، خوانش به صورت مرحله‌بندی شده و بر اساس همه انواع شناسایی‌شده به هاپلوبلاک‌ها تبدیل شدند. هر هاپلوبلاک یک امتداد توالی کاملاً فازی را توصیف می‌کند که امکان توصیف کامل آن ناحیه را فراهم می‌کند که نشان‌دهنده یک آلل مادری یا پدری است. قابل ذکر است که ۷۱.۲ درصد از ژنوم را می‌توان به ۱۶۱۹۳ هاپلو بلوک با طول کل هاپلو بلوک ۲.۳ میلیارد جفت باز و اندازه هاپلو بلوک میانه ۴۰۳۰۲ جفت باز (محدوده: ۲.۹-۱ میلیون جفت باز) تبدیل کرد. یک تمایز واضح در اندازه هاپلوبلاک بین مناطق بین ژنی (میانگین ۱۴۹۶۰ جفت باز) و ویژگی‌های ژن کد

جدول ۱: عملکرد جستجو واریانت‌های فارماکوژن ها

جستجوی واریانت‌ها	SNVs			Indels		
	دقت (٪)	بازیابی (٪)	F1 (٪)	دقت (٪)	بازیابی (٪)	F1 (٪)
هاپلوتایپ جستجوی GATK	۹۰.۱	۸۶.۱۲	۹۴.۴۷	۹۹.۹۲	۹۹.۹۶	۹۹.۸۸
DeepVariant (مدل CCS)	۹۸.۳۷	۹۸	۹۸.۷۴	۹۹.۹۲	۱۰۰	۹۹.۸۴



شکل ۲: وضوح هاپلوبلاک ویژگی‌های GENCODE. A توزیع طول هاپلوبلاک طبقه بندی شده توسط ویژگی‌های ژنکد و مناطق بین ژنی، همپوشانی با فارماکوژن‌ها با رنگ قرمز برجسته شده است. B برای هر ویژگی کدکننده پروتئین، درصدی که در مقایسه با طول ویژگی به هاپلوبلاک‌ها تفکیک شد. خط قرمز میانگین طول خوانش را نشان می‌دهد. اکثر هاپلوبلاک‌ها بزرگ‌تر از میانگین طول خوانش هستند، که نشان می‌دهد نه تنها طول خوانش، بلکه تعداد انواع هتروزیگوت برای طول یک هاپلوبلاک تعیین‌کننده است.



شکل ۳: پیچیدگی فارماکوژن‌ها و نسبت حل شده در هاپلو بلوک‌ها در (A) ، فارماکوژن‌ها و پیچیدگی آنها به درصد پوشش داده شده در هاپلو بلوک‌ها مربوط می‌شود. در ژن‌های پررنگ موجود در پاسپورت فارماکوژنومیک در همه جا (B .U-PGx) برای ژن‌های موجود در CPIC دستورالعمل‌های DPWG، تعداد دستورالعمل‌های عملی موجود با درصد هر ژنی که به هاپلوک‌ها تبدیل می‌شود، ترسیم می‌شود. اقدام پذیر به عنوان دستورالعملی تعریف می‌شود که تغییر دوز یا تغییر دارو را توصیه می‌کند. برای هر ژن، درصد حل شده در هاپلوبلاک‌ها در هدر پانل گنجانده شده است. کنسرسیوم اجرای فارماکوژنتیک بالینی CPIC، گروه کاری فارماکوژنتیک هلندی DPWG.

(میانگین ۵۶۷۴۳ جفت باز)، شکل ۲A مشاهده شد. اکثریت قریب به اتفاق ویژگی‌های Gencode به طور کامل به هاپلوبلاک‌ها تبدیل شدند. به طور خاص، ۷۱ درصد از تمام ویژگی‌های کدکننده پروتئین می‌توانند کاملاً مرحله‌بندی شوند (۹۰ درصد) و ۲۲ درصد دیگر تا حدی مرحله‌بندی شده‌اند در حالی که ۷ درصد حل‌نشده باقی مانده‌اند (10≤ درصد مرحله‌ای). الگوهای مشابهی برای سایر ویژگی‌های Gencode مشاهده شد. به نظر نمی رسد طول خوانش عامل محدود کننده اصلی در حل هاپلوتیپ‌ها باشد زیرا درصد یک ویژگی تحت پوشش در هاپلوبلوک‌ها مستقل از طول ویژگی است (شکل ۲B). علاوه بر این، اکثر هاپلوبلاک‌ها (۵۷.۷٪) از میانگین طول خوانش فراتر می‌روند، که نشان می‌دهد نه طول خوانش، بلکه تعداد واریانت‌های هتروزیگوت و تعداد خوانده شده در یک منطقه ژنومی مشخص، عوامل محدود کننده در ساخت هاپلوبلاک هستند.

فارماکوژن ها

برای هر یک از ۱۰۰ فارماکوژن انتخاب شده، بخشی از ژن‌های واقع در یک منطقه پیچیده تعیین شد - با مجموعه‌ای که به عنوان مناطق ژنومی تعریف می‌شود که با تکرارهای قطعه‌ای (SD) یا تکرار همپوشانی دارند. در مجموع، ۱۵ فارماکوژن به عنوان ۱۰۰٪ پیچیده طبقه بندی شدند در حالی که هشت فارماکوژن هیچ همپوشانی با SDs یا تکرار نشان ندادند (شکل ۳A).

برای هر یک از ۱۰۰ جایگاه، تقریباً همه انواع را می‌توان با دقت نام برد (دقت و یادآوری ۹۹.۸٪ <). فازبندی بعدی منجر به هاپلوبلاک‌هایی با طول متوسط ۱۴۰۴۷۳ جفت باز شد که در نتیجه اکثریت (۷۳/۱۰۰) ویژگی‌ها به طور کامل به هاپلوبلاک تبدیل شدند (شکل ۳A). مهم‌تر از همه، از ۱۵ فارماکوژن طبقه‌بندی‌شده به‌عنوان کاملاً پیچیده، ۹ فارماکوژن می‌توانند کاملاً فازبندی شوند، ۴ فارماکوژن حداقل ۶۰ درصد و دو مورد آخر نمی‌توانند فازبندی شوند. از ژن‌های پیچیده ۳۵، HLA مورد از ۳۷ به طور کامل برطرف شدند، دو مورد باقی مانده (HLA-DRB5 و HLA-DRB1) به ترتیب ۶.۴٪ و ۶۷.۱٪ برطرف شدند.

با این وجود، چندین فارماکوژن مهم تنها می‌توانند تا حدی به هاپلوبلاک تبدیل شوند. به عنوان مثال، G6PD، DPYD و CYP2C19 به ترتیب ۰، ۵۵ و ۳۴ درصد حل شدند. از آنجایی که G6PD روی کروموزوم X قرار دارد و فرد توالی‌یابی‌شده مذکر است، نمی‌توان مکان را به دو آلل تبدیل کرد و در نتیجه ۰ درصد مکان در هاپلوبوک‌های فازی پوشانده می‌شود. برای DPYD علت در ترکیبی از طول ژن طولانی (۹۰۰۰۰۰~ جفت باز) و تعداد کمی از انواع است که منجر به کشش‌های بزرگ بدون واریانت‌های هتروزیگوت می‌شود که منجر به شکسته شدن هاپلوبوک‌ها می‌شود. برای CYP2C19، بخش بزرگی در مرکز ژن وجود دارد که برای همه انواع هموزیگوت است. به طور خاص، در کل مکان CYP2C19، ۵۲ واریانت وجود دارد که ۳۳ واریانت آن هموزیگوت هستند که منجر به بلوک‌های فازی تکه تکه می‌شود. با این حال، همانطور که همه مناطق توالی‌یابی شده اند، هنوز هم می‌توان با استفاده از دستورالعمل‌ها و فرضیات مرحله‌بندی گروه کاری فعلی فارماکوژنتیک هلندی (DPWG) و کنسرسیوم اجرای فارماکوژنتیک بالینی (CPIC) هاپلوتیپ‌ها و فنوتیپ‌ها را اختصاص داد.

برای ارزیابی سودمندی بالینی، دیپلوتیپ‌ها و فنوتیپ‌ها بر اساس پانل واریانتی از کنسرسیوم فارماکوژنومیکس Ubiquitous (U-PGx) و یک خط لوله که قبلا توسعه یافته بود، اختصاص داده شدند. در مجموع ۱۴۱۸ واریانت در ۱۰ فارماکوژن کلیدی موجود در پانل شناسایی شد که از این تعداد ۳۸ واریانت در پانل فنوتیپینگ در نظر گرفته شد. انواع مرتبط بالینی در ژن‌های CYP3A5، CYP2D6 و VKORC1 شناسایی شدند. برای CYP3A5، واریانت rs776746 (g.99672916 C > T) روی هر دو آلل یافت شد که منجر به یک ژنوتیپ CYP3A5*3/*3 و یک فنوتیپ متابولیزر ضعیف شد. برای CYP3A5 وضعیت PM غیرقابل اجرا در نظر گرفته می‌شود، زیرا این فنوتیپ رایج‌ترین فنوتیپ در قفقازی‌ها است. برای CYP2D6 و VKORC1 فنوتیپ استنباط شده از واریانت وحشی متفاوت بود. در جایگاه CYP2D6، هر دو واریانت (g.42128945 C > T) rs3892097 و rs1065852 (g.42130692 G > A) هتروزیگوت بودند. با فازبندی، مشخص شد که واریانت‌ها روی همان آلل قرار دارند که منجر به دیپلوتیپ CYP2D6*1/*4 و فنوتیپ متابولایزر میانی (IM) CYP2D6 است. علاوه بر این، با توجه به حضور شبه CYP2D7 غیر عملکردی که بیش از

٪۹۰ از توالی خود را با CYP2D6 به اشتراک می‌گذارد، مهم است که هر واریانت تداخل خوانش CYP2D7 را برای تعیین دقیق هاپلوتیپ‌های CYP2D6 حذف کنیم. قرائت‌ها به اندازه کافی طولانی بودند تا تمایز واضحی بین CYP2D6 و CYP2D7 بدون هیچ گونه قرائت نقشه برداری مبهم وجود داشته باشد. همین امر برای CYP2B6 و CYP2B7P کاذب آن و برای جایگاه CYP3A که همه ژن‌ها همسانی توالی بالا را به اشتراک می‌گذارند، مشاهده شد. برای VKORC1، یک واریانت هموزیگوت (NC_000016.10: g.31093557 G > A) شناسایی شد که منجر به ژنوتیپ 1173TT و منجر به کاهش فعالیت شد. به طور کلی، این نتایج نشان می‌دهد که طبق دستورالعمل‌های اجماع عمومی در دسترس، این فرد نیاز به تنظیم دوز برای داروهایی دارد که سوبسترای CYP2D6 و VKORC1 هستند.

ارتباط بالینی

در مجموع، ۱۵ ژن موجود در این مطالعه در دستورالعمل‌های CPIC و/یا DPWG نشان داده شده‌اند که در مجموع ۵۶ و ۶۷ برهمکنش ژن-دارویی به ترتیب برای دستورالعمل‌های DPWG و CPIC ایجاد می‌شود (شکل ۳). از این تعداد ۱۰ ژن (۷/۶۶ درصد) به طور کامل در هاپلو بلوک‌های فازی حل شدند. ژن‌هایی که به طور کامل برطرف شدند در ۳۵ برهمکنش ژن-دارو در DPWG و ۳۵ برهمکنش ژن-دارو در CPIC نقش دارند. برای ژن‌های باقی‌مانده، واریته‌ها هنوز می‌توانند با دقت شناسایی شوند، که براساس عملکرد بالینی فعلی که از داده‌های ژنتیکی غیر فازی استفاده می‌کند، امکان تعیین هاپلوتیپ را فراهم می‌کند.

بحث

در این مطالعه ما نشان داده‌ایم که توالی‌یابی طولانی، جستجو با کیفیت بالا را در تمام فارماکوژن‌های منتخب ارائه می‌دهد. در مقایسه با تجزیه و تحلیل گسترده ژنوم، نتایج برای ژن‌های PGx با توجه به دقت جستجو و وضوح هاپلو بلوک‌های فازی بزرگ‌تر برتر است. علاوه بر این، اکثر فارماکوژن‌های انتخاب شده می‌توانند به طور کامل در هاپلو بلوک‌های مرحله‌ای حل شوند.

بر اساس جستجو واریانت به تنهایی، داده‌های کل ژنوم خوانده شده می‌تواند برای PGx معمولی مشابه روش استفاده از NGS استفاده شود. علاوه بر این، توالی‌یابی طولانی مزیت حل آلل‌های پدری و مادری را ارائه می‌دهد. با توجه به ماهیت چندشکلی فارماکوژن‌ها، احتمال اینکه یک فرد دارای چندین واریانت در یک فارماکوژن باشد بسیار زیاد است و اهمیت فازبندی هاپلوتیپ را افزایش می‌دهد. علاوه بر این، این فراوانی بالای واریانت‌ها منجر به هاپلوبلاک‌های بزرگ‌تر برای فارماکوژن‌ها در مقایسه با ویژگی‌های Gencode شد.

توالی‌یابی طولانی از نظر تشخیص SNV با توالی‌یابی کوتاه قابل مقایسه است و از نظر فازبندی هاپلوتیپ و SVهای پیچیده بهتر عمل می‌کند. فازبندی هاپلوتیپ به طور بالقوه می‌تواند تفاوت بین فنوتیپ متابولایزر میانی استنباط شده (دو واریانت کوتاه کننده در یک آلل) و فنوتیپ متابولایزر ضعیف (دو واریانت کوتاه کننده در آلل‌های مختلف) ایجاد کند. استراتژی‌های هاپلوتایپینگ فعلی PGx از فازبندی محاسباتی استفاده می‌کنند که منجر به مرحله‌بندی دقیق در مقیاس جمعیتی می‌شود. از آنجایی که تنظیمات دارو در سطح فردی انجام می‌شود، دقت در مورد مرحله‌بندی برای یک فرد بسیار مهم است. در اینجا نشان داده‌ایم که توالی‌یابی طولانی مدت اکثر فارماکوژن‌ها را قادر می‌سازد تا بدون نیاز به داده‌های شجره‌نامه یا فازبندی محاسباتی، به‌طور کامل به هاپلوبوک‌ها تبدیل شوند.

توالی‌یابی طولانی همچنین توصیف کاملی از هر واریانت در مکان‌های انتخابی PGx، از جمله واریانت‌ها ساختاری و نادر ارائه می‌دهد، همانطور که با دقت و یادآوری بالا برای SNVها، Indels و SVها نشان داده شده است. به عنوان مثال، میانگین طول خوانده شده (۱۳.۴kbp) تقریباً سه برابر بزرگتر از اندازه مکان CYP2D6 (4.4kbp) است که امکان توصیف کامل مکان و CNVهای بالقوه را فراهم می‌کند. تفاوت بزرگ بین DeepVariant و GATK برای Indels را می‌توان با استفاده از داده‌های طولانی خوانده شده PacBio CCS برای آموزش تماس گیرنده DeepVariant توضیح داد. GATK با حالت خطای توالی‌یابی کوتاه به عنوان پایه، با ۱۰۰ برابر جایگزینی بیشتر از ایندل طراحی شد. از طرف دیگر DeepVariant حالت خطا را از داده‌های آموزشی PacBio HiFi یاد گرفته است که نسبت ایندل ۳۰ برابر بیشتر از جایگزینی دارد. به طور خاص، شناسایی تکرارهای Indels و پشت سر هم با استفاده از خوانش طولانی و DeepVariant به طور قابل توجهی بهبود یافته است. این تفاوت یک بار دیگر مزیت افزوده خوانش طولانی را نسبت به توالی‌یابی کوتاه در رابطه با شناسایی واریانت پیچیده برجسته می‌کند.

برای فرد مورد مطالعه، ۱۴۱۸ SNV در جایگاه‌های PGx بالینی انتخاب شده (۱۰ ژن) شناسایی شد که ۹۴ درصد آن‌ها کاملاً فازی بودند، که نشان‌دهنده فراوانی واریانت‌ها در فارماکوژن‌ها است. علاوه بر این، ماهیت مرحله‌ای این داده‌ها می‌تواند به بهبود درک ما از هاپلوتیپ‌ها و ترکیب‌های مختلف کمک کند. بنابراین، فناوری‌های توالی‌یابی طولانی مدت پتانسیل تغییر دانش ما در مورد عوامل ژنتیکی را دارند که در پاسخ دارویی متغیر نقش دارند.

قبل از اجرای توالی‌یابی طولانی در عمل بالینی، ابزارهایی برای کمک به تفسیر مورد نیاز است. چندین گروه برای توسعه چنین ابزارهای ترجمه‌ای برای PGx تلاش کرده اند. با این حال، هنوز محدودیت‌هایی برای این ابزار وجود دارد. اولا، آنها اغلب طیف وسیعی از واریانت‌های شناخته شده و هاپلوتیپ‌های مرتبط با آنها را پوشش می‌دهند. با این حال، برای هر *-هاپلوتیپ تأثیر بالینی مشخص نیست، بنابراین گاهی اوقات منجر به هاپلوتایپی می‌شود که اثر آن ناشناخته است و اجرای آن را در عمل بالینی دشوار می‌کند. ثانیا، ابزارها همیشه نتیجه یکسانی را برای یک فرد ارائه نمی‌دهند، که نشان می‌دهد مفروضاتی که این ابزارها بر اساس آن‌ها هستند قابل مقایسه نیستند. برای اینکه فقط هاپلوتیپ*های مرتبط بالینی را در تجزیه و تحلیل خود لحاظ کنیم، تحلیل خود را از ابزار بالینی به پانل واریانت‌ها تعریف شده توسط کنسرسیوم U-PGx محدود کرده ایم. با این حال، باید توجه داشت که این امر منجر به حذف اکثر واریانت‌ها در همه جایگاه‌های PGx می‌شود، به دلیل این واقعیت که هنوز دانش کافی در مورد عملکرد این واریانت‌ها وجود ندارد.

برای نشان دادن تأثیر خوانش طولانی بر PGx بالینی، نتایج توالی‌یابی را در چارچوب دستورالعمل‌های DPWG و CPIC ارزیابی کرده‌ایم. بر اساس واریانت‌های ژنتیکی مشاهده شده در فرد مورد مطالعه، دستورالعمل‌ها تنظیم دارو یا دوز را برای ۲۲ دارو توصیه می‌کند. از بین تمام تداخلات ژن-دارو در دستورالعمل‌ها (۵۳ برای DPWG و ۵۴ برای CPIC) اکثریت قریب به اتفاق (۳۵ برای هر دو) با یک ژن پیچیده (جزئی) مرتبط بود که می‌تواند به طور کامل در یک هاپلوبلاک حل شود. همانطور که در این مطالعه نشان دادیم، توالی‌یابی طولانی قادر به حل این پیچیدگی‌ها و ساختن هاپلو بلوک‌های بزرگ است که امکان جستجو هاپلوتیپ دقیق‌تری را فراهم می‌کند. پانل‌های SNV و توالی‌یابی کوتاه، از سوی دیگر، قادر به شناسایی دقیق وتریانت‌ها هستند، اما در توانایی آنها برای حل همه پیچیدگی‌ها و در رابطه با فازبندی هاپلوتیپ محدود هستند.

با این وجود، لازم به ذکر است که همه فارماکوژن‌ها را نمی توان به طور کامل برطرف کرد. دلیل اصلی این امر فقدان واریانت‌های هتروزیگوت برای امکان ساخت هاپلوبلوک بود. این به نوبه خود منجر به شکسته شدن هاپلوبوک‌ها و فارماکوژن‌ها می‌شود که نمی توانند به طور کامل برطرف شوند. برای فردی که مطالعه کردیم، این اثر به ویژه برای CYP2C19 و DPYD آشکار بود. با این حال، شناسایی واریانت هنوز در کل مکان ژنی امکان پذیر بود که امکان تعیین هاپلوتیپ غیر فازی را فراهم می‌کرد. برای این ژن‌هایی که نمی‌توانند به طور کامل حل شوند، روش‌های هاپلوتیپ مرسوم مبتنی بر داده‌های توالی‌یابی غیر فازی همچنان می‌توانند به کار روند که

منجر به پیش‌بینی هاپلوتیپ و فنوتیپ مطابق با عملکرد بالینی فعلی می‌شود. علاوه بر این، برای DPYD، سه واریانت از چهار واریانت مرتبط بالینی هنوز مرحله‌بندی شده بودند، که دو تای آن‌ها در هاپلوبوک یکسان بودند. نشان‌دهنده عدم فازبندی کامل به این معنا نیست که هیچ یک از انواع مربوطه نمی‌تواند مرحله‌بندی شود. از آنجایی که پوشش در تمام فارماکوژن‌ها کافی بود، این عدم فازبندی به دلیل ساختار ژنتیکی فرد ایجاد می‌شود، به دلیل فقدان واریانت‌های هتروزیگوت در این ناحیه، و نه با توالی‌یابی به خودی خود، این به راحتی قابل حل نیست. برای فرد دیگری، همین مشکل هاپلوبوک‌های شکسته ممکن است در ژن‌های دیگر بسته به ژنتیک آنها مشاهده شود. در حالی که توالی‌یابی طولانی مدت برای فارماکوژنومیک بالینی امیدوارکننده به نظر می‌رسد، هزینه‌ها و زمان چرخش مرتبط با آن در حال حاضر برای تشخیص PGx با توان بالقوه بالا بسیار زیاد است. در حال حاضر، این باعث می‌شود توالی‌یابی طولانی با آرایه‌های SNV سریع مورد استفاده در PGx بالینی سازگار نباشد. با این حال، هزینه‌های توالی به سرعت در حال کاهش است. علاوه بر این، ژنوتیپ پیشگیرانه محبوب‌تر می‌شود که باعث می‌شود زمان چرخش طولانی‌تر دیگر مسئله‌ای نباشد.

در این مطالعه از داده‌های ژنتیکی یک نمونه DNA با کیفیت بالا استفاده شد. در عمل بالینی، کیفیت بالا ممکن است همیشه تضمین نشود. با این وجود، کاربردهای قبلی توالی‌یابی طولانی مدت در یک محیط بالینی یا با استفاده از DNA به دست آمده بالینی منجر به نتایج با کیفیت خوبی شده است. علاوه بر این، از سال ۲۰۲۰ یک گردش کار ورودی DNA بسیار کم PacBio که تنها به ۵ اینچ از DNA نیاز دارد، در دسترس بوده است. بنابراین انتظار می‌رود که نتایج توالی‌یابی با کیفیت بالا را بتوان با نمونه‌های بالینی جمع‌آوری شده به طور معمول به دست آورد.

دقت و ارزش توالی‌یابی طولانی قبلاً در کل داده‌های ژنوم بررسی شده است، که ممکن است یک رویکرد هدفمند همانطور که در اینجا ارائه کردیم غیر ضروری به نظر برسد. با این حال، به خوبی ثابت شده است که پیچیدگی نواحی فارماکوژنومیک ژنوم، سنجش‌های فعلی را در حل ساختار ژنتیکی آن‌ها به خطر می‌اندازد و در نتیجه قابلیت اطمینان و کامل بودن سنجش‌های فنوتیپی را محدود می‌کند. تفاوت در ساختار ژنتیکی فارماکوژنتیک در مقایسه با ژن‌های کدکننده پروتئین عمومی، برون‌یابی مستقیم از نتایج کل ژنوم را غیرقابل اعتماد می‌کند. مهمتر از همه، آنها حاوی واریانت‌های بیشتری هستند که با هم بر پاسخ دارویی تأثیر می‌گذارند. این تعداد زیاد پلی‌مورفیسم‌ها منجر به این فرضیه می‌شود که فارماکوژن‌ها به دلیل فراوانی بیشتر واریانت‌های هتروزیگوت، آسان‌تر فازبندی می‌شوند، همانطور که در مطالعه ما تأیید شد. در واقع، دقت در فارماکوژن‌ها بیشتر از سایر ژن‌ها بود، در حالی که خواندن کوتاه دقت بسیار پایین‌تری در تشخیص واریانت‌های ژنتیکی در این مناطق پیچیده دارد. توانایی توالی‌یابی طولانی مدت برای حل فارماکوژن‌ها قبلاً در مطالعات توالی‌یابی هدفمند نشان داده شده بود. با این حال، این مطالعه با هدف ارائه یک نمای کلی جامع از کاربرد توالی‌یابی طولانی مدت در حل فارماکوژن‌های پیچیده و اطلاع‌رسانی در مورد مناطقی بود که همچنان چالش برانگیز هستند.

این مطالعه به داده‌های با کیفیت بالا از یک موضوع محدود شد و به عنوان اثبات مفهومی برای استفاده از توالی خوانی طولانی در PGx عمل کرد. علیرغم این محدودیت، ما احساس می‌کنیم که این به عنوان یک مطالعه اثبات مفهوم برای بررسی پتانسیل توالی خوانی طولانی برای PGx کافی است. بر اساس این داده‌ها در مورد دقت فراخوانی و توانایی تجزیه فارماکوژن پیچیده به هاپلوبوک‌های فازی، نتیجه می‌گیریم که داده‌های توالی‌یابی طولانی مدت فرصت‌های خوبی برای روشن کردن مکان‌های پیچیده PGx و فازبندی هاپلوتیپ ارائه می‌دهد در حالی که فراخوانی دقیق واریانت در فارماکوژن‌های انتخابی حفظ می‌شود.

منبع:

https: //www.nature.com/articles/s41397-021-00259-z