استفاده از توالی یابی طولانی برای روشن کردن مسائل اساسی و پیچیده فارماکوژنومیک

نیوشا دهرویه[1]

۱- کارشناسی ارشد ژنتیک ، دانشگاه آزاد، تهران، ایران
پژوهشگر مرکز تحقیقات پزشکی شخصی آمیتیس ژن

خلاصه

استفاده از فارماکوژنومیک در عمل بالینی در حال تبدیل شدن به استاندارد مراقبت است. با این حال، به دلیل ترکیب ژنتیکی پیچیده فارماکوژن‌ها، در حال حاضر همه واریانت‌های ژنتیکی در نظر گرفته نشده است. در اینجا، ما کاربرد توالی‌یابی طولانی را برای حل مسائل فارماکوژن‌های پیچیده با آنالیز یک نمونه با مشخصه نشان می‌دهیم. این داده‌ها از خواندن‌های طولانی تشکیل شده است که برای حل هاپلوک‌های فازی پردازش شده‌اند. ۷۳ درصد از فارماکوژن‌ها به طور کامل در یک هاپلو بلوک فازی پوشانده شدند، از جمله ۱۵/۹ ژن که ۱۰۰ درصد پیچیده هستند. دقت فراخوانی متغیر در فارماکوژن‌ها بالا بود، با فراخوانی ٪۹۹.۸ و دقت ٪۱۰۰ برای SNVها و ٪۹۸.۷ دقت و ٪۹۸.۰ برای ایندلز. برای اکثر تداخلات ژن-دارو در دستورالعمل‌های DPWG و CPIC، ژن‌های مرتبط می‌توانند به طور کامل حل شوند (به ترتیب ٪۶۲ و ٪۶۳). این یافته‌ها با هم نشان می‌دهند که داده‌های توالی‌یابی طولانی مدت فرصت‌های امیدوارکننده‌ای را برای روشن کردن فارماکوژن‌های پیچیده و فازبندی هاپلوتیپ ارائه می‌کنند و در عین حال فراخوانی دقیق را حفظ می‌کنند.

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، همه واریانت‌های ژنتیکی در پاسخ به دارو و فعالیت آنزیمی را نمی‌توان با سنجش‌های ژنتیکی معمول PGx، به دلیل عوامل متعدد، توضیح داد. اول، ژنوتایپینگ فعلی قادر به حل کامل ساختار ژنتیکی همه ژن‌های دخیل در پاسخ دارویی نیستند. دوم، مکانیسم اثر دارو و یا مسیر متابولیکی آن همیشه به طور کامل شناخته نشده

اســت. برای ارزیابی اینکه چه بخشی از واریانت ژنتیکی است و چه بخشی را می‌توان توسط عوامل دیگر توضیح داد، ضروری اســت که بتوانیم تمام مولفه‌های ژنتیکی را که باعث پاســخ دارویی متغیر می‌شوند توضیح دهیم. با این حال، این مورد به چالش کشــیده می‌شود زیرا اکثر فارماکوژن‌ها حداقل تــا حدی در مناطق پیچیده ژنومی قرار دارند یا دارای انواعی مانند تکرارهای پشــت سر هم و ترکیبــات هیبریدی شــبه ژن هســتند. فن‌آوری‌های ژنوتایپینگ در حال حاضر بر اســاس ریزآرایه‌های SNV (Single Nucleotide Variant) یــا توالی‌یابــی کوتــاه خوانده می‌شــوند. هر دو رویکرد در توصیف این مناطق پیچیده محدود هســتند، زیرا آنها قــادر به حل و فصل مناسب و قابل اعتماد مناطق بسیار همولوگ و شناسایی واریانت‌هــای PGx نیســتند. علاوه بر ایــن، با فازبندی هاپلوتیپ می‌توان تعیین کرد که آیا وریته‌ها روی یک آلل قرار دارند یا روی آلل‌های مختلف قرار دارند، که به طور بالقوه منجر به تفاوت‌هایی در انتســاب فنوتیپ می‌شود. در حال حاضر، دیپلوتیپ‌های PGx بر اساس عدم تعادل پیوندی مرحله بندی می‌شــوند. در حالی که این منجر به هاپلوتیپ‌های دقیق در مقیاس جمعیتی می‌شــود، اما همیشه به فرضیات دقیق در سطح فردی منجر نمی‌شود. تأثیر این چالش‌ها در عمل بالینی زیاد اســت. به عنوان مثــال، ژن پیچیده CYP2D6 در متابولیســم ۲۰ تا ۳۰ درصد داروهای معمول تجویز می‌شــود و نمی‌توان آن را به طور کامل با توالی خوانی کوتاه مشخص کرد.

در ســال‌های اخیر، فناوری‌های توالی‌یابی طولانی مدت از Oxford Nanopore و PacBio نشان داده‌اند که قادر به توصیف مناطق پیچیده ژنومی (دارویی) هستند. برای این مناطق، خواندن طولانی و با کیفیت بالا به طور قابل توجهی دقت فراخوانی انواع را بهبود می‌بخشــد و امکان تفکیک دیپلوتیپ‌های کاملاً فازی را فراهم می‌کند.

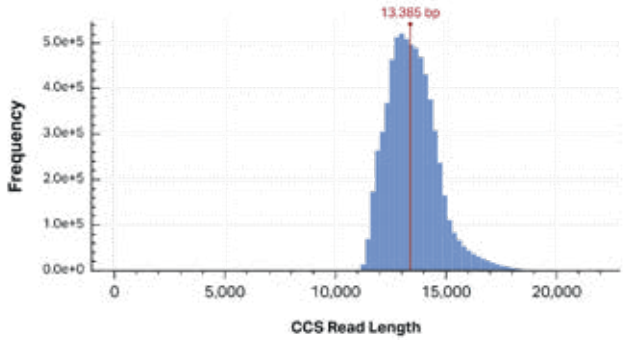
ارزش توالی‌یابــی طولانی مدت برای اهداف تشــخیصی بیماری قبلاً نشــان داده شــده است. نشــان داده شده اســت که توالی‌یابی PacBio می‌توانــد CYP2D6 را با پوشــش کل مکان ژن در یک خواندن طولانی با کیفیت بالا مشــخص کند. اخیراً، توالی‌یابــی طولانی نیز برای ژن‌های HLA در رابطه با PGx اعمال شــده است. علاوه بر این، کاربرد آن در بســیاری از سنجش‌های تحقیقاتی تشخیصی بالینی چالش برانگیز مانند تکرارهای طولانی مدت در ژن FMR1 مرتبط با ســندرم X شکننده و در حل ژن PKD1 برای شناســایی جهش‌هــای مرتبط با بیماری کلیه پلی‌کیستیک استفاده شده است. در نهایت، توالی‌یابی طولانی، مرحله بندی هاپلوتیپ را بدون نیاز به رویکردهای محاسباتی و یا اطلاعات خانوادگی تسهیل می‌کند. ایــن می‌تواند در PGx اهمیت حیاتی داشــته باشــد که منجر به پیش‌بینی دقیق‌تر فنوتیپ می‌شــود. ترکیب پیچیدگی PGx و مرحله بندی هاپلوتیپ نشــان می‌دهــد که توالی خوانی طولانی مدت پتانســیل بهبود قابل ملاحظه‌ای را در توانایی ما برای پیش بینی صحیح فنوتیپ‌های متابولیزکننده دارو دارد. در این مقاله اثبات مفهوم، ما پتانســیل توالی‌یابی طولانی مدت PacBio را برای حــل نواحی پیچیده PGx با اســتفاده از داده‌های توالی‌یابی موجود از ژنوم به خوبی مشخص‌شده در نمونه مرجع بطری HG002 (GIAB) ارزیابی می‌کنیم.

نتایج

توضیحات داده‌ها

داده‌هــای توالی‌یابــی قبلــی منتشــر شــده از نمونه HG002 GIAB به خوبی مشــخص شده به دست آمد. ایــن داده شــامل ۶,۷۲۸,۱۲۳ خوانش با طــول میانه ۱۳.۴ kb اســت که ۹۷.۵ درصد از ژنــوم را با میانگین پوشــش نقشــه برداری ۲۸ برابر پوشش می‌دهد (شکل ۱). تقریبــاً ۵ میلیــون واریانــت ژنتیکی با اســتفاده از GATK (Genome Analysis Toolkit) HaplotypeCaller و DeepVariant شناسایی شد.

شــکل ۱: خوانش توزیع طول. توزیع خوانــدن طول ژنوم در نمونه بطری HG002 پس از توالی‌یابی بر روی پلت فرم توالی Pacific Bioscience و ساخت توالی اجماع دایره‌ای.



دقت و بررسی با دقت بالا در جستجو واریانت‌ها

برای ۱۰۰ فارماکوژن انتخاب شده، دقت و یادآوری در مقایسه با مجموعه حقیقت معیار GIAB v3.3 تعیین شد. برای SNVها، GATK HaplotypeCaller و DeepVariant به دقت و فراخوانی مشابهی بالای ۹۹.۸ درصد دست یافتند (جدول ۱). با این حال، بررسی DeepVariant عملکرد بسیار بهتری در تشخیص ایندل (<۹۸٪) در مقایسه با GATK (دقت: ۹۴.۵٪ و یادآوری: ۸۶.۱٪) به دست آورد. هنگام مقایسه با نتایج گسترده ژنوم گزارش شده توسط ونگر دقت و یادآوری در تشخیص انواع در فارماکوژن‌ها برتر است. هنگام طبقه بندی نتایج در مناطق پیچیده، دقت بالا باقی ماند، با یادآوری و دقت بیش از ۹۵٪ برای همه مناطق برای Indels و SNVs. برای بررسی GATK، دقت کمتر بود (۱۰۰٪-۸۵ در مقایسه با ۱۰۰٪-۹۷ برای بررسی DeepVariant). کاهش دقت 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 تنظیم 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 بندی شده در مجموعه معیار شناسایی شد. دو نتیجه منفی کاذب و یک نتیجه مثبت کاذب در نظر گرفته شد. با هم، ارزیابی عملکرد تشخیص SVs در مناطق PGx منجر به یافتن ۹۱.۳٪ و دقت ۹۴.۵٪ شد. یادآوری و دقت بالا در فارماکوژن‌ها نشان می‌دهد که با استفاده از داده‌های توالی خوانی طولانی در مقایسه با معیارهای فعلی، دقت از دست نمی‌رود، در حالی که تشخیص واریانت‌های پیچیده ژنتیکی را بهبود می‌بخشد.

فازبندی هاپلوتیپ و هاپلوبلاک‌ها

با استفاده از واتس‌هپ، خواندن‌ها به صورت مرحله‌بندی شده و بر اساس همه واریانت‌های شناسایی‌شده به هاپلوبلاک‌ها تبدیل شدند. هر هاپلوبلاک یک امتداد توالی کاملاً مرحله‌ای را توصیف می‌کند که امکان توصیف کامل آن ناحیه را فراهم می‌کند که نشان‌دهنده یک آلل مادری یا پدری است. قابل ذکر است، 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 ژنکد (میانگین ۵۶۷۴۳ جفت باز)، شکل 2A مشاهده شد. اکثریت قریب به اتفاق ویژگی‌های Gencode به طور کامل به هاپلوبلاک‌ها تبدیل شدند. به طور خاص، ۷۱٪ از همه ویژگی‌های کدگذاری پروتئین می‌توانند کاملاً مرحله بندی شوند (۹۰٪ ≤) و ۲۲٪ اضافی تا حدی مرحله بندی شدند در حالی که ۷٪ حل نشده باقی ماندند (۱۰٪ ≥ فازی). الگوهای مشابهی برای سایر ویژگی‌های ژنکد مشاهده شد. به نظر نمی‌رسد طول خواندن عامل محدود کننده اصلی در حل هاپلوتیپ‌ها باشد زیرا درصد یک ویژگی تحت پوشش در هاپلوبلوک‌ها مستقل از طول ویژگی است (شکل 2B). علاوه بر این، اکثر هاپلوبلاک‌ها (۵۷.۷٪) از میانگین طول

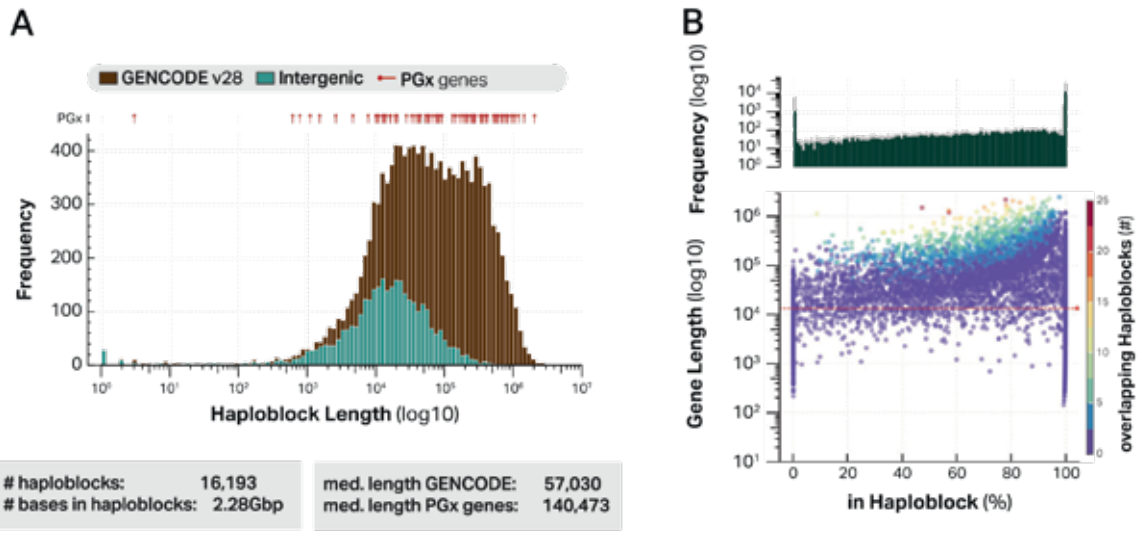
جدول ۱ عملکرد جستجوی واریانت‌ها برای فارماکوژن‌ها.

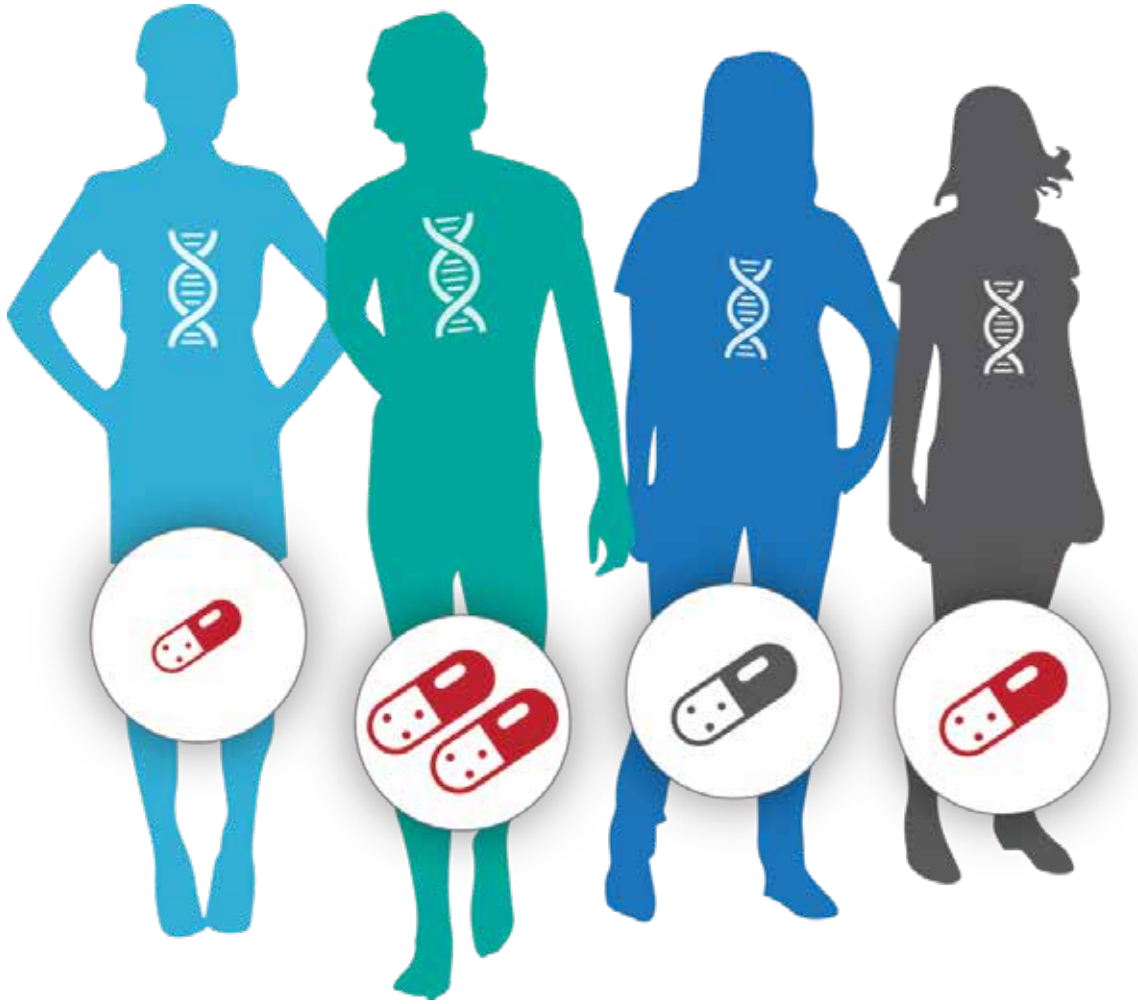
جستجوی واریانت	SNVs			Indels		
	دقت، درستی (٪)	جستجو (٪)	F۱ (٪)	دقت، درستی (٪)	جستجو (٪)	F۱ (٪)
جستجوی هاپلوتیپ GATK	۹۹.۸۸	۹۹.۹۶	۹۹.۹۲	۴۷..۹۴	۸۶.۱۲	۹۰.۱۰
DeepVariant (مدل CCS)	۹۹.۸۴	۱۰۰.۰	۹۹.۹۲	۹۸.۷۴	۹۸.۰۰	۹۸.۳۷

با معیار GIAB نسخه ۳.۳.۲ اندازه‌گیری شد. با استفاده از جستجوگر واریانت GATK و DeepVariant. واریانت تک نوکلئوتیدی SNV، درج و حذف Indels، جعبه ابزار آنالیز ژنومی GATK، توالی اجماع دایره‌ای CCS.

شکل ۲: وضوح هاپلوبلاک ویژگی های GENCODE.

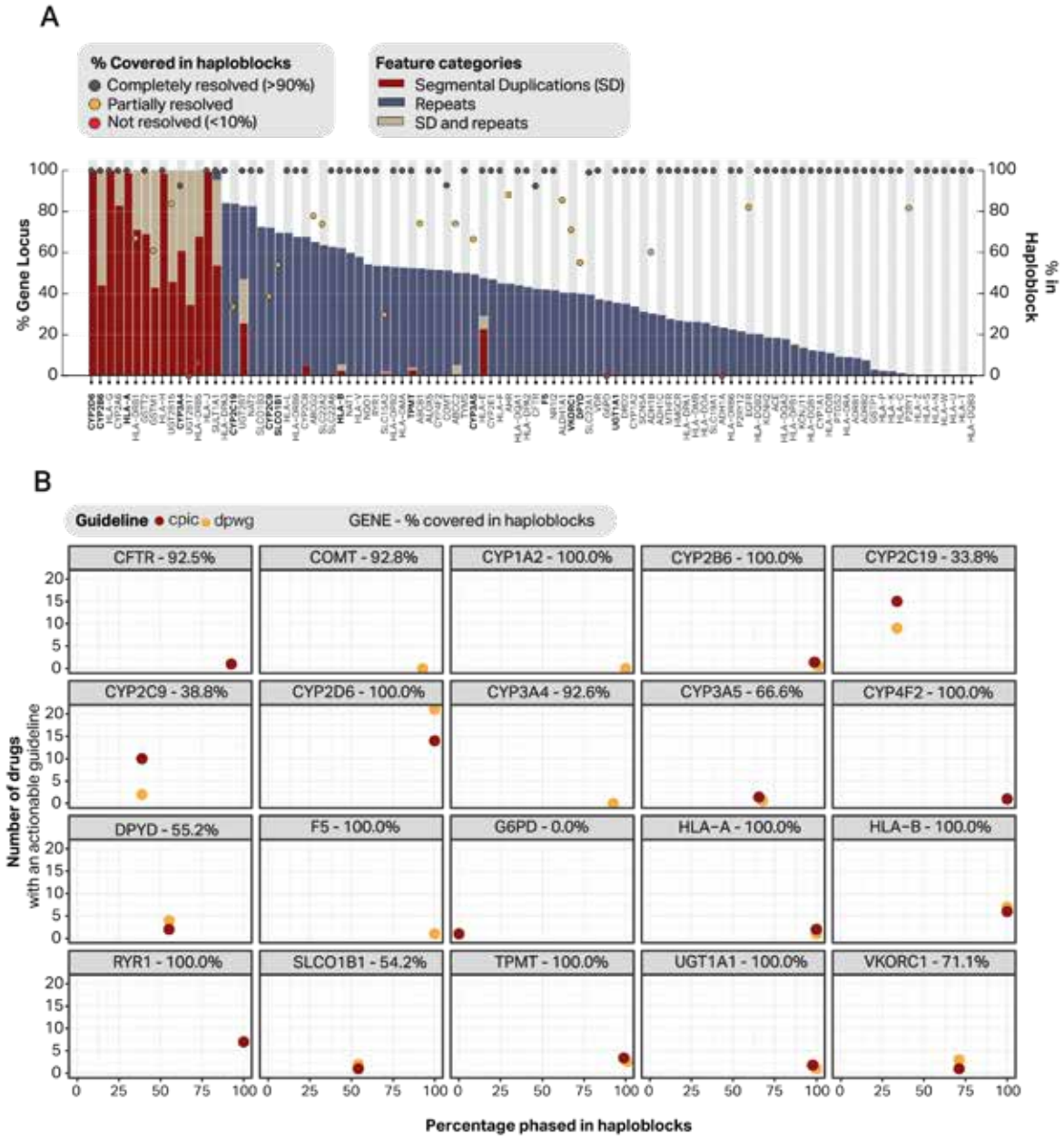
A: توزیع طول هاپلو بلوک طبقه بندی شـــده توســـط ویژگی‌های ژنکد و مناطق بین ژنی، همپوشـــانی با فارماکوژن ها با رنگ قرمز برجســـته شـــده است. B: برای هر ویژگی کدکننده پروتئین، درصدی که در مقایسه با طول ویژگی به هاپلوبلاک‌ها تفکیک شد. خط قرمز میانگین طول خوانش را نشـــان می‌دهد. اکثر هاپلوبلاک‌ها بزرگ‌تر از میانگین طول خواندن هستند، که نشان می‌دهد نه طول خواندن، بلکه تعداد انواع هتروزیگوت برای طول یک هاپلوبلاک تعیین‌کننده است.





شکل ۳: پیچیدگی فارماکوژن ها و نسبت حل شده در هاپلو بلوک ها.

در (A)، فارماکوژن‌ها و پیچیدگی آنها به درصد پوشـش داده شـده در هاپلوبلاک‌ها مربوط می شـود. در ژن‌های پررنگ موجود در پاسـپورت فارماکوژنومیک فراگیر (U-PGx). B برای ژن‌های موجود در CPIC دستورالعمل‌های DPWG، تعداد دستورالعمل‌های عملی موجود با درصد هر ژنی که به‌صورت فازی به هاپلوکک تبدیل می‌شـود، ترسـیم می‌شـود. اقدام پذیر به عنوان دستورالعملی تعریف می شـود که تغییر دوز یا تغییر دارو را توصیه می‌کند. برای هر ژن درصد حل شـده در هاپلوبلاک ها در هدر پانل گنجانده شده است. کنسرسیوم اجرای فارماکوژنتیک بالینی CPIC، گروه کاری فارماکوژنتیک هلندی DPWG.



خواندن فراتر می‌روند، که نشان می‌دهد نه طول خواندن، بلکه تعداد گونه‌های هتروزیگوت و تعداد خوانده‌های هم تراز با یک منطقه ژنومی معین، 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 که با مجموعه‌ای که به عنوان مناطق ژنومی تعریف می‌شود و 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 لوکوس، تقریباً همه واریانت‌ها را می‌توان با دقت جستجو کرد (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 به‌طور کامل فازبندی شوند، ۴ تا برای حداقل ۶۰ درصد و دو فارماکوژن آخر نمی‌توانند فازبندی شوند. از ژن‌های پیچیده HLA، ۳۵ مورد از ۳۷ ژن کاملاً برطرف شدند، دو مورد باقی مانده (HLA-DRB5 و HLA-DRB1) به ترتیب ۶.۴ و ۶۷.۱ درصد 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 بر روی کروموزوم X قرار دارد و توالی‌یابی شده مذکر است، نمی‌توان مکان را به دو آلل تبدیل کرد و در نتیجه ٪۰ از لوکوس در هاپلوبوک‌های فازی پوشانده می‌شود. برای DPYD علت در ترکیبی از طول ژن طولانی (۹۰۰۰۰۰~ جفت باز) و تعداد کم واریانتی است که منجر به کشش‌های بزرگ بدون واریانت‌های هتروزیگوت می‌شود که منجر به هاپلوبوک‌های شکسته می‌شود. برای CYP2C19، بخش بزرگی در مرکز ژن وجود دارد که برای همه واریانت‌ها هموزیگوت است. به طور خاص، در کل لوکوس CYP2C19 حدود ۵۲ واریانت نوع دارد که ۳۳ وریته‌ی آن هموزیگوت هستند که منجر به بلوک‌های فازی تکه تکه می‌شود. با این حال، از آنجایی که همه مناطق توالی‌یابی شده‌اند، هنوز امکان اختصاص هاپلوتیپ‌ها و فنوتیپ‌ها با استفاده از دستورالعمل‌ها و فرضیات مرحله بندی گروه کاری فعلی فارماکوژنتیک هلندی (DPWG) و کنسرسیوم اجرای فارماکوژنتیک بالینی (CPIC) وجود دارد.

برای ارزیابی کاربرد بالینی، دیپلوتیپ‌ها و فنوتیپ‌ها بر اساس پانل واریانت 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 میان ۳۸ وریته در پانل فنوتیپ در نظر گرفته شد. واریانت‌های مرتبط بالینی در ژن‌های CYP2D6، CYP3A5 و VKORC1 شناسایی شدند. برای CYP3A5، وریته‌ی 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 عمل تلقی می‌شود زیرا این فنوتیپ رایج‌ترین فنوتیپ در قفقازی‌ها است. برای CYP2D6 و VKORC1 فنوتیپ استنباط شده از وریته‌ی وحشی متفاوت بود. در لوکوس CYP2D6، هر دو وریته rs3892097 (g.42128945 C > T) 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 وجود شبه CYP2D7 غیر عملکردی که بیش از ٪۹۰ از توالی آن را با CYP2D6 به اشتراک می‌گذارد، حذف هرگونه تداخل خوانش CYP2D7 برای تعیین دقیق هاپلوتیپ‌های CYP2D6 بسیار مهم است. قرائت‌ها به اندازه کافی طولانی بودند تا تمایز واضحی بین CYP2D6 و CYP2D7 بدون هیچ گونه قرائت‌های نقشه برداری مبهم وجود داشته باشد. همین امر برای CYP2B6 و شبه CYP2B7P آن و برای جایگاه CYP3A که همه ژن‌ها همسانی توالی بالا را دارند 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 نشـــان می‌دهد که، طبق دستورالعمل‌های اجماع عمومی در دســـترس، این فرد نیاز بـــه تنظیم دوز برای داروهایی دارد که سوبســـترای CYP2D6 و VKORC1 هستند.

ارتباط بالینی

در مجمـــوع، ۱۵ ژن موجـــود در ایـــن مطالعـــه در دستورالعمل‌های CPIC و یا DPWG نشان داده شده‌اند که منجر بـــه ۵۶ و ۶۷ تعامـــل ژن-دارو به ترتیب برای دستورالعمل‌های DPWG و CPIC می‌شود (شکل ۳). از این تعـــداد ۱۰ ژن (۷/۶۶ درصد) به طور کامل در هاپلو بلوک‌های فازی حل شدند. ژن‌هایی که به طور کامل حل شـــدند در ۳۵ تا از برهمکنش‌های ژن-دارو در DPWG و بـــرای ۳۵ برهمکنـــش ژن-دارو در CPIC نقش دارند. برای ژن‌های باقی‌مانده، هنـــوز هم می‌توان واریانت‌ها را به دقت شناسایی کرد و با توجه به عملکرد بالینی فعلی که از داده‌های ژنتیکی غیر مرحله‌ای اســـتفاده می‌کند، تخصیص هاپلوتیپ را ممکن می‌سازد.

بحث

در این مطالعه اثبات مفهوم، نشان داده‌ایم که توالی‌خوانی طولانی مدت جســـتجوهایی با کیفیت بالایی را در تمام فارماکوژن‌های انتخابی به دســـت می‌دهد. در مقایسه با آنالیزگســـترده ژنوم، نتایج برای ژن‌های PGx با توجه به دقت فراخوانـــی و وضوح هاپلوک‌های فازی بزرگ‌تر برتر اســـت. علاوه بر این، اکثر فارماکوژن‌های انتخاب شـــده می‌تواننـــد به طور کامل در هاپلـــو بلوک‌های فازی حل شوند.

بر اســـاس فراخوانی نوع به تنهایـــی، داده‌های کل ژنوم طولانـــی خوانده شـــده را می‌توان بـــرای PGx معمولی مشابه روش اســـتفاده از NGS اســـتفاده کرد. علاوه بر این، توالی‌یابی طولانـــی مدت مزیت حل آلل‌های پدری و مادری را ارائه می‌دهد. با توجه به ماهیت چندشـــکلی فارماکوژن‌هـــا، احتمـــال اینکه یک فـــرد دارای چندین گونه در یک فارماکوژن باشـــد بسیار زیاد است و اهمیت فازبنـــدی هاپلوتیپ را افزایش می‌دهد. علاوه بر این، این فراوانی زیاد انواع منجـــر به هاپلوبلاک‌های قابل توجهی بزرگ‌تر برای فارماکوژن‌ها در مقایســـه بـــا ویژگی‌های Gencode شد.

توالی‌یابی طولانی از نظر تشخیص SNV با توالی خوانی کوتاه قابل مقایســـه اســـت و از نظر فازبندی هاپلوتیپ و SVهای پیچیده بهتـــر عمل می‌کند. فازبندی هاپلوتیپ به طور بالقـــوه می‌تواند تفاوت بیـــن فنوتیپ متابولایزر میانی استنباط شـــده (دو نوع کوتاه کننده در یک آلل) و فنوتیـــپ متابولایزر ضعیف (دو نوع کوتـــاه کننده در آلل‌های مختلف) ایجاد کند. استراتژی‌های هاپلوتایپینگ فعلی PGx از مرحله‌بندی محاســـباتی استفاده می‌کنند که منجر به مرحله‌بندی دقیق در مقیاس جمعیتی، اما نه همیشه در سطح فردی می‌شود. از آنجایی که تنظیمات دارو در سطح فردی انجام می‌شود، دقت در مورد مرحله بندی برای یک فرد بســـیار مهم اســـت. در اینجا نشان داده‌ایـــم که توالی‌یابی طولانی مـــدت اکثر فارماکوژن‌ها را قادر می‌ســـازد تا بدون نیاز به داده‌های شـــجره‌نامه یا فازبندی محاســـباتی، به‌طور کامـــل در هاپلوبوک‌ها قرار بگیرند.

توالی‌یابی طولانی همچنین توصیف کاملی از هر واریانت در لوکوس‌هـــای انتخابـــی PGx، از جملـــه واریانت‌های ســـاختاری و نادر ارائه می‌دهد، همانطـــور که با دقت و یادآوری بـــالا برای SNVها، Indels و SVها مشـــخص می‌شـــود. به عنـــوان مثال، میانگین طول خوانده شـــده (۱۳.۴kbp) تقریبـــاً ســـه برابـــر بزرگتر از انـــدازه منبع CYP2D6 (4.4kbp) اســـت، که امـــکان توصیف کامل لوکوس و CNVهای بالقوه را فراهم می‌کند. تفاوت بزرگ بیـــن DeepVariant و GATK برای Indels را می‌توان با استفاده از داده‌های طولانی خوانده شده PacBio CCS برای آمـــوزش جســـتجوگر DeepVariant توضیح داد. GATK بـــا حالت خطای توالی خوانـــدن کوتاه به عنوان پایه، با ۱۰۰ برابر جایگزینی بیشتر از ایندل طراحی شد. از طرف دیگـــر 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 ســـر هم با استفاده از خوانش طولانی و DeepVariant به طور قابل توجهی بهبود یافته است. این تفاوت مزیت افزوده خواندن طولانی نسبت به توالی‌یابی کوتاه را برای شناسایی واریانت‌های پیچیده بار دیگر برجسته می‌کند.

برای فرد مـــورد مطالعـــه، ۱۴۱۸ SNV در لوکوس‌های PGx بالینی انتخاب شده (۱۰ ژن) شناسایی شد که ۹۴

درصد آن‌ها کاملاً فازبندی شـــده بودند، که نشان‌دهنده فراوانـــی واریانت‌ها در فارماکوژن‌ها اســـت. علاوه بر این، ماهیت مرحله‌ای ایـــن 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 پاســخ دارویی متغیر نقش دارند.

قبـــل از اجـــرای توالی‌یابـــی طولانی در عمـــل بالینی، ابزارهایی برای کمک به تفســـیر مورد نیاز است. چندین گروه برای توســـعه چنین ابزارهای ترجمه‌ای برای PGx تلاش کرده‌اند. با این حـــال، هنوز محدودیت‌هایی برای ایـــن ابزار وجود دارد. اول، آن‌ها اغلب طیف وســـیعی از واریانت‌های شناخته شده و هاپلوتیپ‌های مرتبط با آن‌ها را پوشـــش می‌دهند. با این حال، بـــرای هر *-هاپلوتیپ تأثیر بالینی مشخص نیست، بنابراین گاهی اوقات منجر به هاپلوتیپی می‌شود که اثر آن ناشناخته است و اجرای آن در عمل بالینی را 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 نیســـتند. برای اینکه فقط *-هاپلوتیپ‌های مرتبـــط بالینی را در آنالیز خود لحاظ کنیم، تحلیل خود را از ابزار بالینی به پانل واریانت‌های تعریف شـــده توسط کنسرســـیوم U-PGx محدود کرده‌ایم. البته لازم به ذکر اســـت که این امر منجر به حذف اکثر واریانت‌ها در همه لوکوس‌های PGx می‌شود، به دلیل این واقعیت که هنوز دانش کافی در مورد عملکرد این واریانت‌ها وجود ندارد.

برای نشان دادن تأثیر خواندن طولانی بر PGx بالینی، ما نتایج توالی‌یابی را در چارچوب دستورالعمل‌های DPWG و CPIC ارزیابی کرده‌ایم. بر اســـاس واریانت‌های ژنتیکی مشاهده شده در فرد مورد مطالعه، دستورالعمل‌ها تنظیم دارو یـــا دوز را بـــرای ۲۲ دارو توصیه می‌کنـــد. از 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) همراه بود که می‌توانست به طور کامل در یک هاپلوبلاک حل شـــود. همانطور که در این مطالعه نشـــان دادیـــم، توالی‌یابی طولانی قادر به حل ایـــن پیچیدگی‌ها و ســـاختن هاپلوبلاک‌های بزرگ اســـت که امکان فراخوانی هاپلوتیپ دقیق‌تری را فراهم می‌کند. پانل‌های SNV و توالی‌یابی کوتاه، از سوی دیگر، قادر به شناسایی دقیق واریانت‌ها هستند، اما در توانایی خود برای حل همـــه پیچیدگی‌ها و در رابطه با فازبندی هاپلوتیپ محدود هستند.

با این وجود، لازم به ذکر اســـت که تمـــام فارماکوژن‌ها نمی‌توانند به طور کامل برطرف شـــوند. دلیل اصلی این امـــر فقدان واریانـــت هتروزیگوت بود تا امکان ســـاخت هاپلوبلـــوک را فراهم کنـــد. این به نوبه خـــود منجر به شکســـته شـــدن هاپلوبوک‌ها و فارماکوژن‌ها می‌شود که به طور کامل قابل حل نیســـتند. برای فردی که مطالعه کردیم، این اثر به ویژه برای CYP2C19 و DPYD آشکار بود. با این حال، شناسایی وریته هنوز در کل لوکوس ژن امکان پذیر بود که امـــکان تعیین هاپلوتیپ غیر فازی را فراهم می‌کرد. برای ایـــن ژن‌هایی که نمی‌توانند به طور کامل حل شـــوند، روش‌های هاپلوتیپ مرسوم مبتنی بر داده‌های توالی‌یابـــی 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 مورد از چهار واریانت مرتبط بالینی هنوز مرحله‌بندی شده‌اند، که دو مورد از آن‌ها در هاپلوبوک یکسان هستند. نشـــان‌دهنده عدم فازبندی کامل به این معنی نیست که هیچ یک از واریانت‌هـــای مربوطه نمی‌تواند مرحله‌بندی شود. از آنجایی که پوشـــش در تمام فارماکوژن‌ها کافی بود، این عدم فازبندی ناشی از ساختار ژنتیکی افراد است، به دلیل فقدان واریانت‌های هتروزیگوت در این ناحیه، و نه با توالی‌یابی به خـــودی خود، این به راحتی قابل حل نیســـت. برای فرد دیگری، همین مشکل هاپلوبوک‌های شکســـته ممکن اســـت در ژن‌های دیگر بسته به ژنتیک آنها مشاهده شـــود. در حالی که توالی‌یابی طولانی برای فارماکوژنومیـــک بالینی امیدوارکننده به نظر می‌رســـد، 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 می‌شـــود توالی‌یابی طولانی با آرایه‌های SNV سریع مورد استفاده در PGx بالینی سازگار نباشد. بـــا این حـــال، هزینه‌های توالی‌یابی به ســـرعت در حال کاهش است. علاوه بر این، ژنوتیپ پیشگیرانه محبوب‌تر می‌شود که باعث می‌شود زمان چرخش طولانی‌تر دیگر مسئله‌ای نباشد.

در این مطالعـــه از داده‌های ژنتیکی یـــک نمونه DNA با کیفیت بالا اســـتفاده شـــد. در عمـــل بالینی، کیفیت

بالا ممکن است همیشه تضمین نشود. با این وجود، کاربردهای قبلی توالی یابی 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 جمع‌آوری‌شده به‌طور معمول به‌دست 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 مناطق فارماکوژنومیک ژنوم، سنجش‌های کنونی 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 فارماکوژن‌ها در مقایسه با ژن‌های کد کننده پروتئین عمومی، برون یابی مستقیم از نتایج کل ژنوم را غیر قابل 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 اطلاع رسانی در مورد مناطقی بود که همچنان چالش برانگیز هستند.

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

مواد و روش ها

توضیحات داده ها

نتایج توالی‌یابی طولانی در دسترس عموم از نمونه GIAB HG002 با توالی‌یابی PacBio تعیین توالی شد و با استفاده از CCS (Circular Consensus Sequencing) خوانده شد. یک نمونه GIAB انتخاب شد زیرا اینها با نتایج معیار موجود بسیار خوب مشخص می‌شوند. خواندن‌های CCS با استفاده از نرم افزار CCS نسخه 3.0.0 ایجاد شد. قرائت‌های HiFi به دست آمده با استفاده از NGMLR aligner v0.2.7 با ژنوم مرجع GRCh۳۸ تراز شد. انواع ژنتیکی با استفاده از GATK HaplotypeCaller (v.4.0.6.0) و DeepVariant (v.0.7.1) شناسایی شدند. مجموعه‌ای از ۶۴ فارماکوژن که قبلاً توسط Lauschke به همراه ژن‌های پیچیده HLA توصیف شده بود برای آنالیز PGx انتخاب شدند.

ساخت هاپلو بلاک

واریانت‌های نامیده شده توسط GATK با استفاده از WhatsHap برای به دست آوردن واریانت‌های SNV فازی و Indel مرحله بندی شدند. از قرائت‌های مرحله‌ای، هاپلوبلاک‌ها ساخته و در فایل‌های GTF و BED ذخیره شدند. هر هاپلوبلاک با تطبیق قرائت‌های مرحله‌ای بر اساس واریانت‌های موجود در آن‌ها ساخته شد تا طول توالی قابل حل را افزایش دهد. یک هاپلوبلاک نشان دهنده یک امتداد توالی ناگسستنی است که بر اساس خوانش‌های مرحله‌ای همپوشانی دارند و زمانی متوقف می‌شود که ناحیه‌ای در ژنوم فقط با خوانش‌های بدون هیچ گونه‌ای پوشانده شود، دیگر تفاوتی در واریانت‌های بین دو آلل وجود ندارد یا اگر منطقه فاقد پوشش باشد. پس از آن، همه جایگاه‌ها به یکی از سه ویژگی طبقه‌بندی شدند: ویژگی‌های ژن کد (v28)، ژن‌های

PGx و ویژگی‌های بین ژنی. جایی که یک ویژگی به عنوان یک منطقه ژنومی مشروح شده مانند ژن‌های کدکننده پروتئین، مناطق تکراری قطعه‌ای، شبه‌ژن‌ها، و غیره تعریف می‌شود. تفسیر مرجع ژن کد برای ویژگی‌های ژنتیکی در ژنوم انسان (نسخه ۲۸) برای بررسی تفکیک‌پذیری هاپلوبلاک لوکوس‌های مهمی مانند ژن‌های کدکننده پروتئین پروژه Gencode با هدف طبقه‌بندی و شناسایی تمام ویژگی‌های ژن در ژنوم انسان از جمله تمام تفسیرها انجام می‌شود. برای هر ژنکد اتوزومال و ویژگی PGx، درصد ویژگی پوشش داده شده در هاپلوبلاک محاسبه می‌شود (تعداد جفت‌های پایه در هاپلو بلوک/طول ویژگی کل). مناطق با پوشش هاپلوبلاک ٪۹۰ به عنوان کاملاً فازی طبقه بندی می‌شوند، در حالی که مناطق بدون هاپلوبلاک همپوشانی بدون فاز باقی می‌مانند. همه مناطق دیگر به عنوان نیمه فازی مشخص شده‌اند.

تکرارهای قطعه‌ای (SD) و ترک‌های تکراری از مرورگر ژنوم UCSC (دانشگاه کالیفرنیا سانتا کروز) به دست می‌آیند. Bedtools برای شناسایی مناطق همپوشانی بین تمام ترک‌ها و فایل‌های تتفسیر مورد بحث استفاده شد. برای هر لوکوس، درصد بخش‌هایی که با SD یا تکرار همپوشانی دارند به عنوان «پیچیده» تعریف می‌شود.

ارتباط بالینی

یک پایپ‌لاین که قبلا توسعه یافته بود برای اختصاص هاپلوتیپ‌ها و فنوتیپ‌ها به فارماکوژن‌های مرتبط بالینی بر اساس دستورالعمل‌های DPWG استفاده شد. ژن‌های انتخاب شده بر اساس پانل کنسرسیوم U-PGx و شامل ۱۰ فارماکوژن کلیدی و ۳۸ وریته بودند. این پایپ‌لاین که در اصل برای داده‌های NGS طراحی شده بود، شامل ژن‌های UGT1A1 و HLA-B که در پنل کنسرسیوم U-PGx وجود دارند به دلیل پیچیدگی‌هایشان نبود. همه ژنوتیپ‌ها بر اساس حضورشان در دستورالعمل‌ها و تعداد داروها با توصیه‌های عملی ارزیابی می‌شوند، جایی که عمل‌پذیر به‌عنوان «تعامل ژن-دارویی که نیاز به تغییر دارو، تنظیم دوز یا نظارت شدید دارد» تعریف می‌شود. برای تمام فارماکوژن‌های ذکر شده در دستورالعمل‌های CPIC و DPWG، تعداد تداخلات ژن-دارو محاسبه می‌شود.

منبع

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