

تنوع موقعیت‌های ژنومی مرتبط با انواع کووید ۱۹ در میان جمعیت‌ها: الگوهای جهانی و ملیتی

سابقه: اطلاعات در مورد توزیع نشانگرهای ژنتیکی مهم بالینی در جمعیت‌های مختلف ممکن است در شرح رویکردهای شخصی برای مدیریت بالینی COVID-19 در غیاب دستورالعمل‌های کلی مفید باشد. هدف: تجزیه و تحلیل فراوانی و الگوهای توزیع دو نشانگر مرتبط با COVID-19 شدید (rs11385942 و rs657152) و جستجوی ارتباط بالقوه بین این نشانگرها و مرگ و میر ناشی از COVID-19 در بین جمعیت روسیه و سراسر جهان.

روش‌ها: ما ژنوتیپ ۱۸۸۳ نمونه از ۹۱ گروه نژادی را که در ۲۸ جمعیت به نمایندگی از روسیه و کشورهای همسایه آن جمع‌آوری شده بودند را مشخص کردیم. همچنین یک مجموعه داده از ۳۲ جمعیت در مناطق دیگر با استفاده از ژنوتیپ‌های استخراج شده یا مستند از پایگاه داده‌های موجود گردآوری کردیم. نقشه‌های جغرافیایی نشان دهنده توزیع فراوانی نشانگرهای آنالیز شده با استفاده از داده‌های به دست آمده ساخته شده‌اند. نتایج: تجزیه و تحلیل نقشه‌ها نشان داد که توزیع rs11385942 از الگوی اوراسیای غربی پیروی می‌کند: این نشانگر در میان جمعیت‌های اروپا، آسیای غربی و آسیای جنوبی شایع است اما در سایر نقاط کره زمین نادر است یا وجود ندارد. قابل توجه است که انتقال از فراوانی بالای rs11385942 به فراوانی پایین در سراسر اوراسیا غیرمنتظره نبوده و از الگوی تغییرات بالینی پیروی می‌کند. توزیع rs657152 همگن‌تر است. تجزیه و تحلیل ارتباط بین فراوانی نشانگرهای مورد مطالعه و ویژگی‌های اپیدمیولوژیک COVID-19 در یک جمعیت نشان داد که فراوانی بیشتر هر دو آلل خطرناک، با مرگ و میر ناشی از این بیماری ارتباط مستقیم دارد.



نجمه شجاعی^۱

۱- کارشناسی علوم آزمایشگاهی، دانشگاه کرمان، ایران
پژوهشگر مرکز تحقیقات پزشکی شخصی آمیتیس ژن

که ارتباط برای کل بخش هاپلوتیپ 5^0 کیلوبایتی نشان داده شد. چنین LD بالایی برای این بخش از کروموزوم ۳ غیرمعمول است و می‌توان آن را با ورود تکاملی اخیر این هاپلوتیپ از نئاندرتال‌ها به مخزن ژن انسان توضیح داد. جایگاه دیگر (rs657152) که ارتباط کمتری با COVID-19 شدید نشان داد، در جایگاه 9q34.2 گروه خونی ABO قرار دارد. طبق مطالعات زبرگ و پابو، فراوانی این SNP در میان جمعیت‌های موجود در پروژه ۱۰۰۰ ژنوم متفاوت است و بیشترین فراوانی در جنوب آسیا (بنگلادش) رخ می‌دهد. زبرگ و پابو همچنین گزارشی ارائه کردند که در آن افراد بنگلادشی الاصل در بریتانیا دو برابر بیشتر از جمعیت عادی در معرض خطر مرگ ناشی از کووید-۱۹ هستند.

بنابراین، ما تصمیم گرفتیم که به بررسی چگونگی این دو SNP مرتبط با COVID-19 شدید در جمعیت‌های مختلف در سراسر جهان با تمرکز بر جمعیت‌های کمتر مورد مطالعه در اوراسیا شمالی توزیع شده‌اند. مجموعه داده‌های مورد استفاده ما شامل ۱۸۸۳ نمونه هستند که نشان دهنده جمعیت اوراسیای شمالی و داده‌های منتشر شده شامل ۳۰۸۸ نمونه از سایر مناطق جهان می‌باشند. هدف ما کشف الگوهای تنوع جغرافیایی این چندشکلی‌ها و بررسی ارتباط احتمالی بین فراوانی آنها و تعداد مرگ و میر و بهبودی کووید-۱۹ در جمعیت‌های مورد مطالعه بود. اطلاعات در مورد توزیع نشانگرهای ژنتیکی مهم بالینی در جمعیت‌های مختلف ممکن است در غیاب دستورالعمل‌های کلی به ایجاد رویکردهای شخصی‌سازی شده برای درمان COVID-19 کمک کند.

هدف اولیه این مطالعه تعیین فراوانی rs11385942 و rs657152 در جمعیت‌های مختلف روسیه که کشوری بسیار ناهمگن از لحاظ قومیتی است، و نزدیک‌ترین همسایگان آن می‌باشد. هدف دوم جستجوی هرگونه ارتباط احتمالی بین فراوانی جمعیت rs11385942 و rs657152 و تعداد موارد کووید-۱۹، بهبودیافتگان و مرگ‌ومیرها بود. هدف دوم جستجوی هرگونه ارتباط احتمالی بین فراوانی جمعیت rs657152 و rs11385942 و تعداد موارد کووید-۱۹، بهبودیافتگان و مرگ‌ومیرها بود.

برای rs657152، ارتباط به صورت ویژه‌ای قوی بود ($r = 0.59, p = 0.02$). این ارتباطات معقول فقط برای مجموعه داده "روسی" مشاهده شد: چنین ارتباطی برای مجموعه داده "جهانی" صادق نیست. این امر را می‌توان به تفاوت در روش مورد استفاده برای جمع آوری آمار COVID-19 در کشورهای مختلف نسبت داد.

جمع بندی: یافته‌های ما نشان می‌دهند که تفاوت‌های ژنتیکی بین جمعیت‌ها سهم کوچک و در عین حال ملموسی در ناهمگونی پاندمی در سراسر جهان دارد. کلمات کلیدی: COVID-19 شدید، نشانگرهای ژنتیکی، rs657152، rs11385942، AB 0 ، جغرافیای ژن

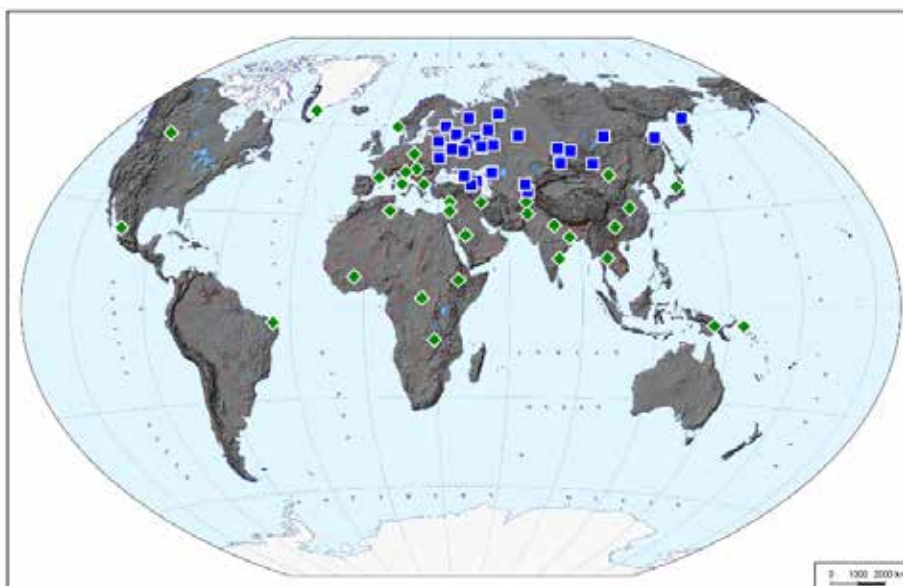
مقدمه

ویروس کرونا جدید که به عنوان سندرم حاد تنفسی ویروس کرونا ۲ شناخته می‌شود (SARS-CoV-2) برای اولین بار در دسامبر ۲۰۱۹ در چین شناسایی شد. این ویروس به سرعت در حال گسترش در سراسر جهان بود و سه ماه بعد سازمان بهداشت جهانی اعلام پاندمی کرد. بیماری ناشی از SARS-CoV-2 به عنوان بیماری ویروس کرونا ۲۰۱۹ (COVID-19) شناخته می‌شود. مسیر بیماری از عفونت بدون علامت تا مرگ ناشی از نارسایی تنفسی متفاوت است. داده‌های اپیدمیولوژیک نشان می‌دهد که عوارض و مرگ و میر ناشی از COVID-19 نه تنها بین افراد در یک جمعیت، بلکه بین کشورها نیز بسیار متفاوت است. در زمان نگارش این مقاله، نرخ روزانه موارد جدید COVID-19 در شمال اروپا (دانمارک، استونی، فنلاند، نروژ) نسبتاً پایین بود، در حالی که اروپای جنوبی (ایتالیا، اسپانیا، فرانسه) از عوارض و مرگ و میر بالاتر رنج می‌بردند. چنین تنوعی ممکن است ناشی از عوامل اجتماعی جمعیت شناختی، محیطی (برای بررسی ارتباط با سطوح ویتامین D به بخش ۵-۳ مراجعه کنید) و عوامل ژنتیکی باشد. مطالعه‌ای در مورد ارتباط گسترده ژنومی، شامل ۱۶۱۰ بیمار و ۲۲۰۵ شرکت‌کننده کنترل از ایتالیا و اسپانیا، دو مکان ژنومی کلیدی مرتبط با نارسایی شدید تنفسی ناشی از COVID-19 را شناسایی کرده است. یکی از آنها (G>GA) rs11385942 در مکان 3p21 است. 3p21، که شامل مجموعه‌ای از شش ژن است که به طور بالقوه مستعد ابتلا به کووید-۱۹ شدید هستند. قابل توجه است

مواد و روش‌ها

مجموعه داده ۱: جمعیت روسیه و کشورهای همسایه آن نمونه‌هایی که نشان‌دهنده جمعیت روسیه و نزدیک‌ترین همسایگان آن در اوراسیای شمالی (مناطق پسا شوروی و مغولستان) توسط Biobank اوراسیای شمالی ارائه شده است. این مطالعه توسط کمیته اخلاق مرکز تحقیقات ژنتیک پزشکی تایید شد. رضایت آگاهانه از هر اهدا کننده اخذ شده است. نمونه‌های آنالیز شده، قلمرو وسیعی را نشان می‌دهد که بخش اعظم روسیه، کشورهای همسایه و اکثر گروه‌های قومی ساکن در این منطقه را در بر می‌گیرد (شکل ۱). مشابه با استراتژی نمونه‌برداری پروژه ۱۰۰۰ ژنوم، مطالعه ما بر روی جمعیت‌های بومی متمرکز بود. به عنوان مثال، جمعیت سیبری شرقی توسط Yakutها و Evenkها، به جز روس‌های قومی که در ۱ تا ۳ قرن گذشته به سیبری نقل مکان کرده بودند، نمایندگی می‌شدند. ما در مجموع ژنوتیپ ۱۸۸۳ نمونه را تعیین کردیم که نشان دهنده ۹۱ جمعیت از ارمنستان، آذربایجان، بلاروس، گرجستان، قزاقستان، قرقیزستان، لیتوانی، مغولستان، روسیه، تاجیکستان، اوکراین و ازبکستان بودند. ما روی دو نشانگر مرتبط با کووید-۱۹ شدید تمرکز کردیم: rs11385942 و rs657152. نمونه‌ها برای rs657152 و rs250 SNP در مجاورت (chr3: 45700000-45900000, GRCh37) با استفاده از کیت BeadChip Infinium Omni5Exome-4 v1.3،

(Illumina ایالات متحده آمریکا)، که بیش از ۴.۵ میلیون SNP را پوشش می‌دهد، تعیین ژنوتیپ شدند. برای تخمین فراوانی نشانگرهای مرتبط با کووید، به نمونه‌ای با اندازه ۱۰۰ کروموزوم در هر جمعیت نیاز داشتیم، اما اکثر جمعیت‌های موجود در مطالعه تنها با چند نمونه نشان داده شدند. بنابراین، نمونه‌هایی از جمعیت‌های نزدیک از نظر جغرافیایی و ژنتیکی گردآوری شدند. شباهت ژنتیکی بین این جمعیت‌ها با استفاده از تعیین ژنوتیپ گسترده ژنومی برای ۴.۵ میلیون SNP موجود در آرایش Illumina تعیین شد. روش ادغام شامل مراحل زیر بود: ۱) فیلتر کردن ورودی و ساده‌سازی با PLINK (geno0.05, maf0.01, mind0.1, indep-pairwise15001500.2) ۲) اجرای PCA با smartpca (۱۱، ۱۰، ۳) شناسایی گروه‌هایی از جمعیت‌هایی که از نظر ژنتیکی همگن بودند اما تنوع بین گروهی مشخصی را نشان دادند. برخی از جمعیت‌ها به دلیل حجم نمونه کوچک و ویژگی‌های ژنتیکی که ادغام را غیرممکن می‌کرد، از تجزیه و تحلیل حذف شدند. روش ادغام با جزئیات بیشتری شرح داده شده است. مجموعه داده نهایی شامل ۱۷۸۵ نمونه از ۲۸ جمعیت ادغام شده بود. متوسط حجم نمونه ۱۲۸ کروموزوم (۶۴ نفر) بود. این جمعیت‌ها، که در زیر به آنها «مجموعه داده روسی» گفته می‌شود، در جدول S1 فهرست شده‌اند. جغرافیای آنها بر روی نقشه نشان داده شده است (شکل ۱).



شکل ۱: جمعیت‌های مورد مطالعه. مربع‌های آبی مکان نمونه‌های جمعیتی را نشان می‌دهد که به طور خاص برای این مطالعه (مجموعه داده روسی) تعیین ژنوتیپ شده‌اند. الماس‌های سبز مکان‌هایی از نمونه‌های جمعیتی را نشان می‌دهند که در منابع دیگر توضیح داده شده و در طول این مطالعه (مجموعه داده‌های جهانی) مجدداً تحلیل شده‌اند.

نزدیک دیگری، برچسب گذاری می‌کند. آرایه SNP مورد استفاده برای تعیین ژنوتیپ مجموعه داده "روسی" بسیار متراکم‌تر (۴.۵ میلیون SNP) از آرایه مورد استفاده قبلی شامل 250 SNP در مجاورت rs113859420 بود. نشانگرهای Ungenotyped با بیگل نسبت داده شد. تعداد تکرارها روی ۲۰۰ تنظیم شد. مجموعه داده پروژه ۱۰۰۰ ژنوم به عنوان مرجع استفاده شد. از آنجایی که آرایه‌های Illumina مورد استفاده برای ژنوتیپ نمونه‌های موجود در مجموعه داده "جهانی" چگالی کمتری (66 SNP در مجاورت rs113859420) داشتند، در ابتدا باید کیفیت انتساب ارزیابی می‌شد. به این منظور، ژنوتیپ نمونه‌های مشابه تولید شده توسط آرایه‌های SNP 250 و SNP 66 را مقایسه کردیم؛ ۱۸۷۱ (۹۹.۴٪) از ۱۸۸۳ ژنوتیپ‌های بین دو مجموعه منطبق شدند، که تأیید می‌کند که انتساب ژنوتیپ‌های rs113859420 از آرایه‌های Illumina با چگالی کمتر کاملاً دقیق بوده است. از این ژنوتیپ‌ها برای شمارش فراوانی آلل rs113859420 در جمعیت‌های مجموعه داده "جهان" استفاده شد (جدول S1).

تحلیل نقشه ای

نقشه‌هایی که توزیع جغرافیایی دو نشانگر مرتبط با کووید را نشان می‌دهند در GeneGeo ترسیم شده‌اند. نقشه‌های اوراسیا شمالی (روسیه و نزدیکترین همسایگان آن) با استفاده از روش شپرد ایجاد شده است. تابع وزن روی ۳ تنظیم شد. شعاع نفوذ ۱۵۰۰ کیلومتر تعیین شد. برای سایر مناطق جهان، نقشه‌های توزیع فراوانی با استفاده از همین روش ساخته شد و با شعاع نفوذ تا ۲۵۰۰ کیلومتر گسترش یافت تا تغییرات فراوانی را به جمعیت‌هایی که در مجموعه داده‌ها گنجانده نشده‌اند مقایسه کند. همچنین تابع وزن را برای هموارسازی بهتر روی ۲ تنظیم کرد تا روند مشاهده شوند.

تحلیل آماری

برای تجزیه و تحلیل نوع ارتباط، به اطلاعاتی در مورد تعداد موارد COVID-19 در هر یک میلیون نفر، تعداد بهبودیافتگان به ازای هر یک میلیون نفر، تعداد مرگ و میر ناشی از COVID-19 در هر یک میلیون نفر، و تعداد مرگ و میر در هر یک از موارد تایید شده COVID-19

مجموعه داده ۲: جمعیت سایر مناطق جهان

علاوه بر مجموعه داده‌ای که جمعیت روسیه را نشان می‌دهد، مجموعه‌ای از جمعیت‌های دیگر مناطق جهان را گردآوری کردیم. از داده‌های ۱۶ مقاله استفاده کردیم که در آن‌ها جمعیت‌های بومی با استفاده از آرایه‌های ژنوتایپینگ گسترده ژنوم توسط Illumina (Illumina700k, Illumina730k, Illumina660k, Illumina650k, Illumina610k, Illumina550k, and Illumina1M) مورد مطالعه قرار گرفتند. مجموعه داده ادغام شده شامل ۳۳۳۶ نمونه تعیین ژنوتیپ شده برای rs11385942 و SNP 66 در مجاورت rs11385942 است. با این مجموعه داده، برخی از نمونه‌ها نیز باید ادغام می‌شدند تا به اندازه نمونه‌های بزرگ‌تر دست یابند. در بیشتر موارد، نمونه‌هایی که نمایانگر همان کشور بودند، جمع‌آوری شده است. در برخی موارد، نمونه‌هایی مربوط به کشورهای هم‌مرز مانند اسپانیا و فرانسه نیز ادغام شدند. دو کشور بزرگ چین و هند، هر کدام با ۳ جمعیت ترکیب شده نشان داده شدند. برای جلوگیری از شمارش مضاعف، نمونه‌های روسیه و کشورهای هم‌مرز آن از این مجموعه داده حذف شدند. مجموعه داده نهایی شامل ۳۲ جمعیت با حجم نمونه متوسط ۱۰۴ کروموزوم بود. این جمعیت‌ها، که به آنها «مجموعه داده‌های جهانی» گفته می‌شود، در جدول S1 فهرست شده‌اند. جغرافیای آنها بر روی نقشه نشان داده شده است (شکل ۱).

انتساب

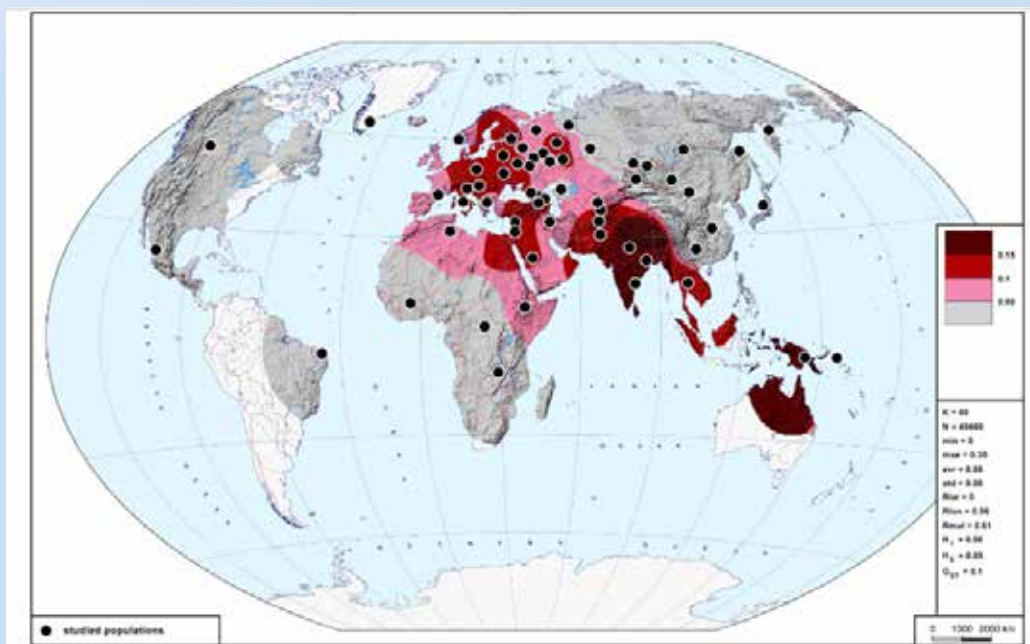
یکی از دو نشانگر مرتبط با کووید ۱۹، rs6۵۷۱۵۲، به طور مستقیم در هر دو مجموعه داده، تعیین ژنوتیپ شد. فرکانس‌های این نشانگر در PLINK محاسبه شده و در جدول S1 نشان داده شده است. نشانگر دوم، rs11385942، در هیچ یک از آرایه‌های ژنوم موجود در حال حاضر، گنجانده نشده است. در طول مطالعه مربوط به ژنوم گسترده که الهام‌بخش کار ما بود، ۷۱۲۱۸۹ SNP تعیین ژنوتیپ شده و ۱۷۰ میلیون SNP استناد شده بودند. ارتباط با COVID-19 شدید برای یکی از SNP‌های منتسب به نام rs11385942 در قوی‌ترین حالت بود. این مطالعه یک بخش هاپلوتایپی تقریباً ۵۰ کیلوبایتی را گزارش کرد که در آن SNP‌ها در LD بالا بودند. rs11385942 این بخش را به اندازه هر SNP

StatSoft Statistica (Dell Statistica, Tulsa, OK, USA) مورد آزمایش قرار گرفت. آستانه معناداری تنظیم شده توسط بونفرونی نیز روی ۰.۰۵ تنظیم شد.

نتایج

تنوع فراوانی rs11385942 در سطح جهانی Rs11385942 قویا با COVID-19 شدید مرتبط است. ما توزیع جغرافیایی انواع آن را در جمعیت های جهان با استفاده از دو مجموعه داده تجزیه و تحلیل کردیم: ۱۸۸۳ نمونه تعیین ژنوتیپ شده که روسیه و کشورهای همسایه آن را نشان می دهد و داده های منتشر شده در مورد ۳۰۸۸ نمونه از مناطق دیگر. مجموعه داده «جهان» حاوی اطلاعاتی درباره فراوانی rs11385942_GA در ۶۰ جمعیت بود (جدول S1). شکل ۲ توزیع جغرافیایی این نشانگر را در سراسر کره زمین نشان می دهد. بیشترین فراوانی آن (۲۰ تا ۳۰ درصد) در جنوب آسیا و پس از آن در غرب آسیا و اروپا (۵ تا ۱۵ درصد) مشاهده می شود. این SNP در جمعیت های شمال آفریقا یافت می شود، اما در شرق آسیا، شمال آسیا (سیبری)، جمعیت های بومی آمریکا و صحرای آفریقا نادر بوده یا شناسایی

نیاز داشتیم. (این اطلاعات آخرین بار در ۱۸ سپتامبر ۲۰۲۰ به روز رسانی شده است). آمار COVID-19 در هر منطقه در دسترس بود و بین جمعیت بومی و غیربومی تبعیضی اعمال نشد. بنابراین، ما ۱۶ منطقه را در روسیه و کشورهای همسایه آن شناسایی کردیم که در آن اکثریت جمعیت را بومیان تشکیل می دادند (به طور متوسط ۸۵٪، طبق آخرین سرشماری) و تجزیه و تحلیل ارتباط را بر روی آن ۱۶ گروه انجام دادیم (جدول S۳). برای تجزیه و تحلیل توزیع نشانگرهای مرتبط با COVID-19 در سطح جهانی، از آن ۱۶ گروه که روسیه و کشورهای همسایه آن و همه جمعیت های موجود در مجموعه داده «جهان» را نشان می دهند، به جز بومیان آمریکا و گرینلند، استفاده کردیم (جدول S۴). تفاوت در فراوانی آلی بین گروه ها با استفاده از آزمون دقیق فیشر و GraphPad InStat (نرم افزار GraphPad، سن دیگو، کالیفرنیا، ایالات متحده آمریکا) مورد آزمایش قرار گرفت. آستانه معناداری تنظیم شده توسط بونفرونی روی ۰.۰۵ تنظیم شد. ارتباط بین فراوانی SNP های مورد مطالعه و پارامترهای اپیدمیولوژیک در هر ۱ میلیون نفر جمعیت با استفاده از ضریب همبستگی پیرسون محاسبه شده در



شکل ۲: تنوع جهانی فرکانس های rs11385942-GA، چهار رنگ مناطقی از ۴ محدوده فراوانی این آلل خطر را مشخص می کنند. نقاط سیاه نشان دهنده جمعیت های آنالیز شده است. اختصارات موجود در شرح آماری موارد زیر را نشان می دهند.

اختصارات: K، تعداد جمعیت های مورد مطالعه. N، تعداد نقاط شبکه نقشه (محاسبه شده بر اساس داده های جمعیت های مورد مطالعه)؛ min، حداقل فراوانی روی نقشه؛ max، حداکثر فراوانی روی نقشه؛ avr، میانگین فرکانس روی نقشه؛ std، انحراف استاندارد؛ Rlat، ضریب همبستگی جزئی بین عرض جغرافیایی و فراوانی در نقشه. Rlon، ضریب همبستگی جزئی بین طول و فراوانی بر روی نقشه. Rmul، ضریب همبستگی چندگانه فراوانی با هر دو طول و عرض جغرافیایی بر روی نقشه.



SNP از یک الگوی تغییرات بالینی پیروی می‌کند، یعنی فراوانی آن به تدریج در طول ۷۰۰۰ کیلومتر، از ۱۳٪ تا ۱۶٪ در اوکراین، بلاروس و غربی‌ترین مناطق روسیه کاهش می‌یابد و در شبه جزیره کامچاتکا و چوکوتکا در سواحل اقیانوس آرام به صفر می‌رسد. متوسط فراوانی rs11385942 در جمعیت‌های ساکن روسیه اروپایی ۱۱٪ در مقابل ۳٪ در سیبری است (جدول S1). در مناطق آسیای مرکزی، فراوانی rs11385942 عموماً پایین است (۱-۴٪)، تاجیکستان یک استثنا قابل توجه است (۱۴٪)، که تفاوت شدیدی با فرکانس‌های rs11385942 در کشورهای همسایه خود دارد. این موضوع با واقعیت نزدیکی جمعیت تاجیکستان از نظر جغرافیایی و ژنتیکی به جمعیت‌های جنوب آسیا، جایی که فراوانی rs11385942 به اوج خود می‌رسد، توضیح داده می‌شود. فراوانی rs11385942 در مناطق جنوب غربی روسیه بسیار شبیه به فراوانی آن در مرز بلاروس و اوکراین بوده و تفاوت از نظر آماری ناچیز است (جدول S1 و S2). در قفقاز، کمترین فراوانی آلل خطر ۶٪ (GA) در مناطق مرکزی (چچن، اینگوشتیا، اوستیای شمالی) مشاهده می‌شود، که به سمت شرق و غرب (۱۰٪) و به ویژه در قفقاز جنوبی (۱۴ درصد) افزایش می‌یابد.

تنوع فرکانس‌های rs657152 در سطح جهانی

شکل ۴ یک نقشه توزیع فرکانس برای rs657152 را نشان می‌دهد. توزیع این نشانگر در مقایسه با rs11385942، همگن‌تر است. تقریباً در تمام جمعیت‌های دنیای قدیم بسیار رایج است. بیشترین فراوانی آن (بیش از ۵۰

نشده است. با توجه به داده‌های اندک کنونی در مورد توزیع rs11385942_GA در آسیای جنوب شرقی و پاپوآ گینه نو، فراوانی این نوع در این دو منطقه افزایش یافته است. مشابه بسیاری از نشانگرهای ژنتیکی دیگر، از جمله DNA میتوکندری و نشانگرهای کروموزومی Y، توزیع جغرافیایی این نشانگر از الگوی شناخته شده "اوراسیای غربی" پیروی می‌کند. جمعیت‌های آسیای جنوبی نزدیکی ژنتیکی بیشتری با اوراسیای غربی نسبت به اوراسیای شرقی دارند. به صورت جالب توجهی، توزیع جغرافیایی rs11385942 با فراوانی‌های بالاتر در جمعیت‌های آسیای جنوبی و فراوانی پایین‌تر در جمعیت‌های غرب آسیا/اروپایی اوراسیای غربی مشخص می‌شود.

برای جبران اجتناب ناپذیری ویژگی شماتیک این توصیف، تجزیه و تحلیل تنوع rs11385942 در سطح جهانی با تجزیه و تحلیل دقیق توزیع جغرافیایی آن در روسیه تکمیل شد.

تنوع فرکانس‌های rs11385942 در سراسر روسیه

مجموعه گسترده از نمونه داده‌های جمع‌آوری شده در روسیه و کشورهای همسایه آن به ما این امکان را می‌دهد که توزیع جغرافیایی rs11385942 را در این کشورها با جزئیات بیشتری تجزیه و تحلیل کنیم. فراوانی این SNP در جدول S1 فهرست شده و بر روی نقشه نمایش داده شده است (شکل ۳). نقشه نشان می‌دهد که هیچگونه گذار ناگهانی از فراوانی بالای rs11385942 در اروپا به فراوانی صفر آن در شمال آسیا وجود ندارد. توزیع این



COVID-19



در زمینه بهبود COVID-19 و نرخ مرگ و میر در آن جمعیت‌ها آنالیز شدند. اطلاعات مربوط به تعداد موارد کووید-۱۹، بهبودیافتگان، مرگ و میرها و تعداد جمعیت در جدول S۳ ارائه شده است. ارتباط منفی ضعیفی (جدول ۱) بین فراوانی هر دو آلل خطر و تعداد مطلق موارد COVID-19، و همچنین تعداد موارد COVID-19 مطابق با اندازه جمعیت ایجاد شد. در مقابل، همبستگی بین هر دو آلل خطر و میزان مرگ و میر، که به عنوان تعداد مرگ و میر در کل موارد COVID-19 محاسبه شد، مثبت بود. ارتباط بین هر دو آلل خطر و نرخ بهبودی محاسبه شده به عنوان تعداد بهبودی مطابق با تعداد کل موارد COVID-19 منفی بود. ارتباط معناداری بین مرگ و میر COVID-19 و فراوانی rs657152 ($r = 0.63$) وجود داشت ($p = 0.01$; جدول ۱)، بر خلاف ارتباط بین مرگ و میر COVID-19 و فراوانی rs11385942 ($r = 0.24$) که از نظر آماری ناچیز ($p = 0.38$) بود. ارتباط بین فراوانی rs657152 و مرگ و میر ناشی از COVID-19 پس از اصلاح بونفرونی به صورت قابل توجه باقی ماند. پس از ادغام ۲ مجموعه داده، ارتباط ایجاد شده ناچیز شد (جدول ۱).

بحث

مطالعه ارتباط گسترده ژنوم توسط Ellinghaus و همکاران یک سیگنال ارتباطی قوی در مکان 3p21.31

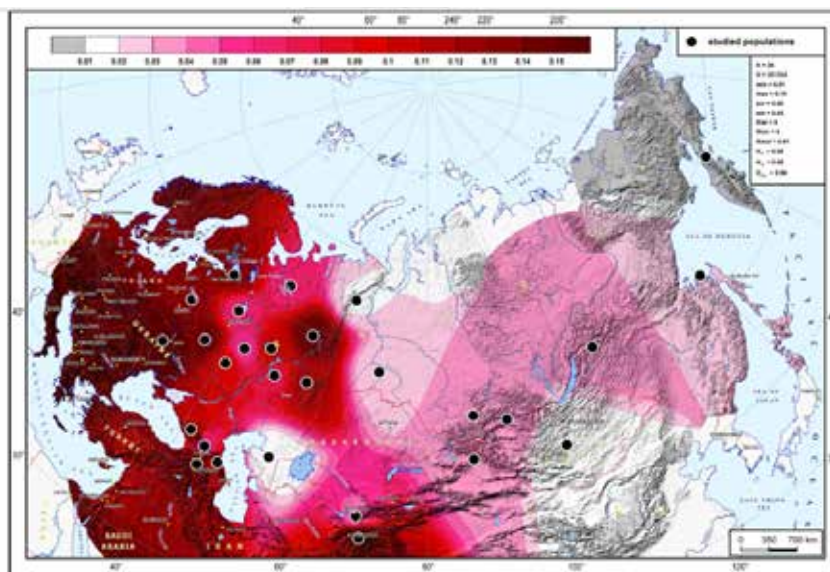
در جنوب صحرای آفریقا مشاهده شده است. در اکثر جمعیت‌های اوراسیا، rs657152 در فرکانس‌های ۴۰ تا ۵۰ درصد رخ می‌دهد. در حاشیه اوراسیا (حاشیه اقیانوس اطلس اروپا، شرق دور، آسیای جنوب شرقی)، فراوانی آن به زیر ۴۰ درصد کاهش می‌یابد. این نشانگر در جمعیت بومی آمریکا و استرالیا تقریباً وجود ندارد.

تنوع فرکانس‌های rs657152 در سراسر روسیه

فراوانی rs657152 در جمعیت روسیه اروپایی از ۳۸٪ تا ۴۲٪ متغیر بوده، قابل مقایسه با بلاروس (۳۳٪، $p > 0.05$) و کمتر از اوکراین است (۵۱٪، $p < 0.05$) (جدول S۱ و S۲). در قفقاز، فراوانی rs657152 در شرق (داغستان، ۵۲٪) بیشترین فراوانی و در غرب (۲۷٪) کمترین فراوانی را دارد. تفاوت در فراوانی rs657152 بین جمعیت این مناطق در مقایسه با سایر مناطق موجود در آنالیز (جدول S2) حائز بیشترین اهمیت است. در آسیا، فراوانی rs657152 در توتان‌ها و مغول‌ها به ۳۸ تا ۳۹ درصد می‌رسد و در جمعیت ازبکستان و قرقیزستان ۵۱ درصد است (تفاوت‌ها در مقایسه با قزاقستان و تاجیکستان، در اینجا ناچیز است؛ جداول S1 و S2).

همبستگی بین فراوانی نشانگرهای مرتبط با

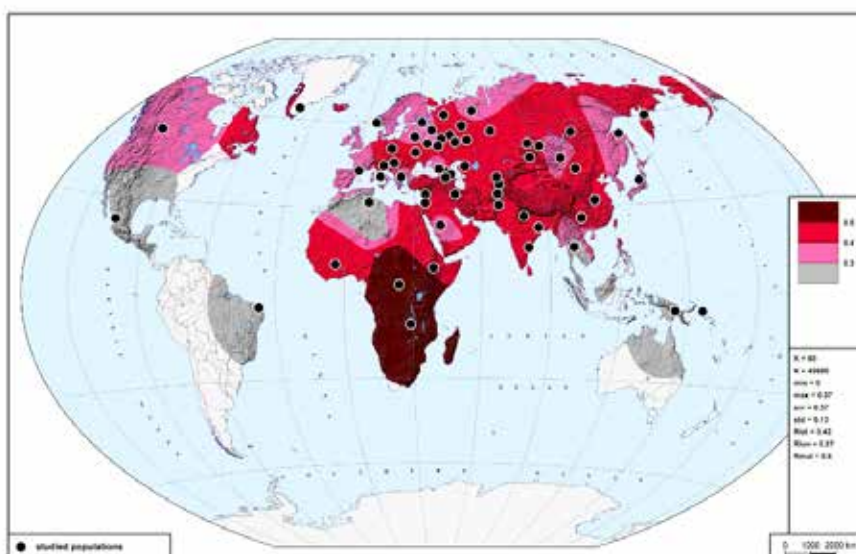
کووید و پارامترهای اپیدمیولوژیک فراوانی هر دو نشانگر در جمعیت‌های مختلف روسیه



شکل ۳: تنوع فرکانس‌های rs11385942_GA در روسیه و کشورهای همسایه آن. طیف فراوانی در اینجا جزئی‌تر از آن چیزی است که در نقشه جهان نشان داده شده است (شکل ۲). نقاط سیاه نشان دهنده جمعیت‌های آنالیز شده است. اختصارات موجود در شرح آماری موارد زیر را نشان می‌دهد. اختصارات: K، تعداد جمعیت‌های مورد مطالعه. N، تعداد گره‌های شبکه نقشه (محاسبه شده بر اساس داده‌های جمعیت‌های مورد مطالعه)؛ min، حداقل فراوانی روی نقشه؛ max، حداکثر فراوانی روی نقشه؛ avr، میانگین فراوانی روی نقشه؛ std، انحراف استاندارد؛ Rlat، ضریب همبستگی جزئی بین عرض جغرافیایی و فراوانی در نقشه. Rlon، ضریب همبستگی جزئی بین طول و فراوانی بر روی نقشه. Rmul، ضریب همبستگی چندگانه فراوانی با هر دو طول و عرض جغرافیایی بر روی نقشه. HT، هتروزیگوسیتی کل. H S، هتروزیگوسیتی درون جمعیت. GST، ضریب تمایز ژنتیکی بین جمعیت‌ها (ضریب G ST Masatoshi Nei).

از مناطق مختلف جهان گردآوری کردیم و دریافتیم که تنوع آن در سطح جهانی از الگوی اوراسیای غربی پیروی می‌کند. مانند بسیاری از چندشکلی‌های دیگر، rs11385942 در اوراسیای غربی (اروپایی‌ها، آسیای غربی و آسیای جنوبی) رایج است، اما در سایر نقاط کره زمین، از جمله جمعیت‌های آسیای شرقی، آسیای شمالی، بومیان آمریکایی و صحرای آفریقا نادر است یا وجود ندارد. (جمعیت‌های آفریقای شمالی مربوط به اوراسیای غربی هستند و rs11385942 را با فراوانی متوسط شامل می‌شوند). این نشانگر، همراه با SNP‌های مرتبط در بخش‌های پلوטיפ به طول ۵۰ کیلوبایت، محصول ترکیب ناندرتال است، که فقدان آن را در صحرای آفریقای جنوبی توضیح می‌دهد که هرگز توسط ناندرتال‌ها مهار نشده است. فقدان این SNP در شرق اوراسیا ممکن است به رانش ژنتیکی پس از واگرایی اوراسیای غربی و شرقی نسبت داده شود. آنالیز دقیق جمعیت روسیه نشان می‌دهد که تفاوت در فراوانی rs11385942 در سراسر شبه قاره‌ی اوراسیا ناگهانی نیست. به عنوان مثال، فراوانی بالای این نشانگر در جمعیت اروپایی به سمت شرق به تدریج کاهش می‌یابد و در سواحل اقیانوس آرام به صفر می‌رسد.

شامل شش ژن (SLC6A20، LZTFL1، CCR9، FYCO1، CXCR6، XCR1) گزارش کرده است: آلل مستعد خطر rs11385942-GA با استعداد ژنتیکی خاص نارسایی حاد تنفسی به دلیل COVID-19 مرتبط بود. فراوانی آلل خطر در بیمارانی که نیاز به تهویه مکانیکی داشتند در مقایسه با بیمارانی که تحت اکسیژن درمانی مکمل بودند، بیشتر بود. از جمله ژن‌هایی که به شدت با COVID-19 مرتبط هستند SLC6A20، LZTFL1 و CXCR6 بودند. ژن SLC6A20 ناقل اسید آمینه سدیم ۱ را رمزگذاری می‌کند که از نزدیک با ACE2، گیرنده ورودی SARS-CoV-۲، تعامل دارد. LZTFL1 عملکرد موبی و حمل و نقل درون فلاژلی را در سلول تنظیم می‌کند. CXCR6 توزیع سلول‌های T خاطره‌ی ساکن بافت را در قسمت‌های مختلف ریه کنترل می‌کند و از یک پاسخ ایمنی پایدار به تهاجم پاتوژن به دستگاه تنفسی اطمینان حاصل می‌کند. آلل خطر GA در کاهش بیان CXCR6 و افزایش بیان SLC6A20 و LZTFL1 در سلول‌های ریه انسان نقش دارد و شدت COVID-19 را تعیین می‌کند. الینگه‌اوس و همکاران متوجه شدند که فراوانی آلل خطر rs11385942 بین جمعیت‌ها متفاوت است، اما الگوی این تنوع را تحلیل نکردند. ما مجموعه‌ای از فرکانس‌های rs11385942 را در جمعیت‌های انسانی



شکل ۴: تنوع جهانی فرکانس‌های rs657152_A. با توجه به مقیاس، چهار رنگ مناطقی با چهار بازه فراوانی این آلل خطر را مشخص می‌کند. نقاط سیاه نشان دهنده جمعیت‌های آنالیز شده هستند. اختصارات موجود در شرح آماری موارد زیر را نشان می‌دهد. اختصارات: K، تعداد جمعیت‌های مورد مطالعه؛ N، تعداد گره‌های شبکه نقشه (محاسبه شده بر اساس داده‌های جمعیت‌های مورد مطالعه)؛ min، حداقل فراوانی روی نقشه؛ max، حداکثر فراوانی روی نقشه؛ avr، میانگین فراوانی روی نقشه؛ std، انحراف استاندارد؛ Rlat، ضریب همبستگی جزئی بین عرض جغرافیایی و فراوانی در نقشه؛ Rlon، ضریب همبستگی جزئی بین طول و فراوانی بر روی نقشه؛ Rmul، ضریب همبستگی چندگانه فراوانی با طول و عرض جغرافیایی روی نقشه است.

جدول ۱: ارتباط بین بهبودی کووید-۱۹، مرگ و میر و فراوانی توزیع نشانگرهای ژنتیکی مورد مطالعه

مجموعه داده "جهانی"		مجموعه داده "روسی"		پارامتر اپیدمیولوژیک تعداد موارد COVID-19 در هر ۱ میلیون نفر جمعیت تعداد بهبودیافتگان به ازای هر ۱ میلیون نفر جمعیت تعداد مرگ و میر به ازای هر یک میلیون نفر جمعیت میزان مرگ و میر (تعداد مرگ و میر به ازای همه موارد تایید شده COVID-19)
rs657152_A	rs11385942_GA	rs657152_A	rs11385942_GA	
(p = 0.44) -0.13	(p = 0.49) 0.11	(p = 0.09) -0.44	(p = 0.50) -0.18	
(p = 0.84) -0.03	(p = 0.43) 0.13	(p = 0.08) -0.46	(p = 0.50) -0.18	
(p = 0.45) -0.12	(p = 0.54) 0.10	(p = 0.89) -0.04	(p = 0.52) -0.17	
(p = 0.38) -0.14	(p = 0.88) -0.03	(p = 0.01) 0.63	(p = 0.38) 0.24	

شکل ۵، حجم وسیع داده‌های مورد استفاده برای ساخت آن است: نقشه دارای ۲۷۵۷ جمعیت بوده که تقریباً ۳۰ برابر بیشتر از تعداد جمعیت موجود در مجموعه داده‌های تحلیل شده در این مطالعه می‌باشد. این نقشه از داده‌های ادبیات مربوط به گروه‌های خونی انباشته شده در قرن بیستم ساخته شده است که ارزش واقعی مجموعه داده‌های قدیمی را نشان می‌دهد: نقشه‌های توزیع فرکانس بر اساس مجموعه داده‌های مدرن (مانند پروژه ۱۰۰۰ ژنوم یا مجموعه داده‌های آنالیز شده در این مطالعه) تا حدودی دقیق هستند، در حالی که داده‌های منتشر شده در نیمه دوم قرن بیستم اطلاعات بسیار دقیق و تصویر قابل اعتمادی را ارائه می‌دهند (شکل ۵). نقشه نشان می‌دهد که بالاترین فراوانی (۳٪) rs657152 و بالاتر) در اروپای غربی رخ می‌دهد. فراوانی rs657152 کمی کمتر در منطقه ولگا-اورال اروپای شرقی مشاهده می‌شود. جمعیت شمال آسیا (سیبری)، آسیای شرقی و آفریقای جنوب صحرا با فراوانی کلی rs657152 (۱۰-۲۰٪) مشخص می‌شود. اکثر جمعیت‌های جنوب آسیا حامل ABO_A در فراوانی متوسط (۱۵-۲۰٪) هستند، اما در جمعیت‌های غرب آسیای مرکزی فراوانی این آلل به اندازه اروپای غربی است.

نتیجه

به طور خلاصه، فراوانی آلل خطر rs11385942 در جنوب آسیا زیاد، در اروپا و غرب آسیا کمی کمتر و در سایر مناطق جهان کم است. فراوانی آلل خطر rs657152 در آفریقا در بالاترین میزان و در اوراسیا متوسط است. گروه خونی A بیشتر در جمعیت‌های اروپایی مشاهده می‌شود و در اوراسیا نسبتاً فراوان است. به عبارت دیگر، فراوانی

سایت ژنومی دیگری که قویاً با COVID-19 شدید مرتبط است rs657152 می‌باشد. این سایت در جایگاه گروه خونی ABO قرار دارد. وجود ارتباط بین سیستم گروه خونی ABO و شدت COVID-19 در شماری از مطالعات گزارش شده است. برای مثال، خطر نارسایی تنفسی مرتبط با COVID-19 در بیماران دارای گروه خونی A در حالت حداکثر بود. در مقابل، بیماران با گروه خونی O در کمترین خطر نارسایی تنفسی مرتبط با COVID-19 قرار داشتند. اگرچه، مطالعات ذکر شده در درجه اول بر خطر عفونت متمرکز بودند اما بر شدت COVID-19 تمرکز نداشتند. اثر محافظتی گروه خونی O به وجود آنتی بادی‌های خنثی کننده علیه گلیکان‌های مرتبط با N نسبت داده شد. همچنین مشخص است که بین جایگاه گروه خونی ABO و بیان فاکتور فون ویلبراند (لوکوس 13p13.31)، که همراه با فاکتور VIII، لخته شدن رگ‌های خونی آسیب دیده را افزایش می‌دهد، ارتباطی وجود دارد. بیماران با گروه خونی غیر O نسبت به بیماران با گروه خونی O، دارای سطوح بالاتر VWF در سلول‌های اندوتلیال ریوی خود هستند که ممکن است نقش سیستم گروه خونی ABO را در COVID-19 توضیح دهد.

قابل ذکر است که rs657152 مستقیماً هیچ گروه خونی ای را رمزگذاری نمی‌کند، اما می‌تواند برای تشخیص انواع نادر ژن ABO مورد استفاده قرار گیرد. انواع "کلاسیک" ژن ABO پیچیدگی پشت این ارتباط را نشان می‌دهد، که ممکن است با تحقیقات آینده شامل توالی یابی کل مکان ABO مشخص شود. در همین حال، بیابید به فراوانی ABO گروه A در جمعیت‌های مختلف نگاه کنیم (شکل ۵). چشمگیرترین نکته در مورد نقشه در

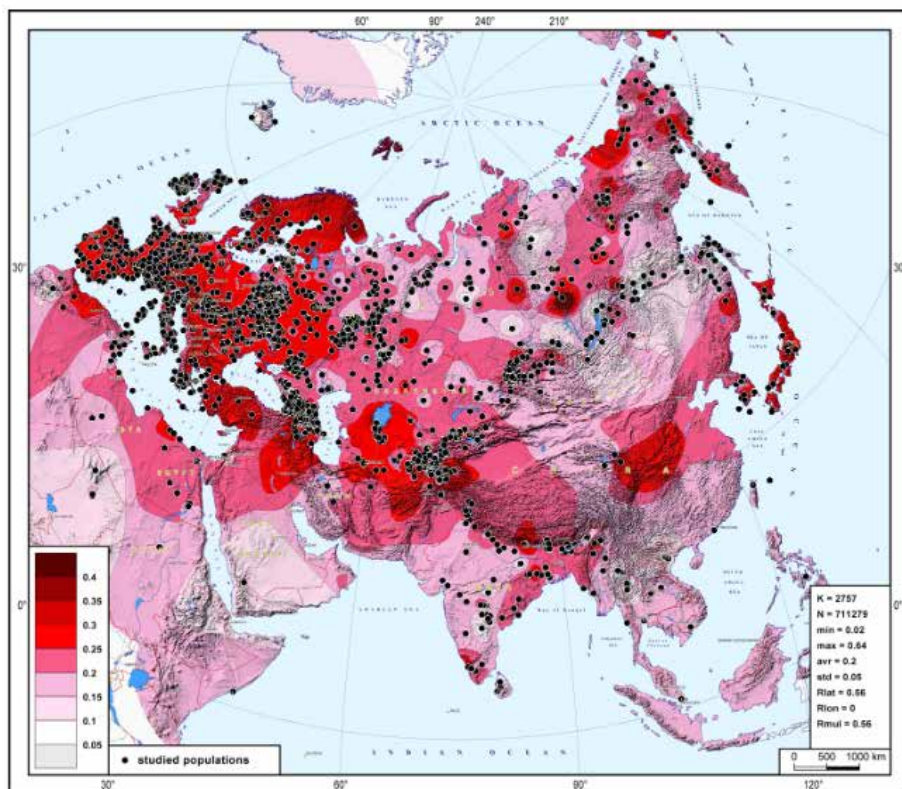
داده «جهانی» به دست نیامده است. این مسئله را می‌توان به تفاوت در روش مورد استفاده برای گزارش موارد غیر شدید در کشورهای مختلف نسبت داد: تعداد موارد گزارشات خفیف و متوسط COVID-19 به طور چشمگیری بر میزان مرگ و میر تأثیر می‌گذارد و تجزیه و تحلیل همبستگی با عوامل ژنتیکی را غیرممکن می‌کند. در مقابل، درون یک کشور، روش نظارت اپیدمیولوژی یکنواخت‌تر است و ارتباط با توزیع فراوانی آلل‌های خطر آشکار می‌شود. اندازه نمونه‌های ما نسبتاً کوچک بودند، بنابراین نوبز آماری ممکن است قدرت ارتباطات ایجاد شده را کاهش داده باشد. ما معتقدیم که همبستگی‌های قوی‌تر بین نرخ بهبود/مرگ و میر کووید-۱۹ و استخر ژنی جمعیت‌های مورد مطالعه در تحقیقات آینده ممکن است آشکار شود و تنوع ژنتیکی بین جمعیت‌ها سهم کوچک و در عین حال ملموسی در ناهمگونی همه‌گیری در بخش‌های مختلف جهان دارد.

منبع:

<https://pubmed.ncbi.nlm.nih.gov/34764675/>

هر یک از آلل‌های خطر تجزیه و تحلیل شده در بین اروپایی‌ها و آسیای جنوبی بالاتر از جمعیت سایر مناطق است.

در حین بررسی اینکه آیا فراوانی بالای آلل‌های خطر مورد مطالعه در یک جمعیت ممکن است وضعیت اپیدمیولوژیک را تشدید کند، یک الگوی منطقی از ارتباطات را کشف کردیم، اگرچه، نمی‌توانیم ادعا کنیم که ارتباط بین این دو باعث ایجاد همبستگی می‌شود. اولاً، ما هیچ ارتباطی بین فراوانی آلل‌های خطر مورد مطالعه و تعداد کل موارد COVID-19 در یک جمعیت پیدا نکردیم، که دور از ذهن نبود زیرا این آلل‌ها بیشتر با سیر شدید بیماری مرتبط هستند تا با حساسیت به عفونت. ثانیاً، همبستگی با پیامدهای COVID-19 مطابق انتظار بود: برای هر دو نشانگر، فراوانی بالاتر آلل‌های خطر با میزان مرگ‌ومیر ارتباط مثبت داشت و روندی با rs11385942 بود. با rs657152 ارتباط قوی ($r = 0.6$) و معنی داری وجود داشت ($p = 0.01$, جدول ۱). این ارتباطات معقول فقط برای مجموعه داده‌های «روسی» مشاهده شد: چنین ارتباطاتی برای مجموعه



شکل ۵: فراوانی توزیع گروه خونی A (سیستم ABO) در جهان. نقشه اصلاح شه از مطالعه قبل می‌باشد. نقاط سیاه نشان دهنده جمعیت آنالیز شده است. اختصارات موجود در شرح آماری موارد زیر را نشان می‌دهند. اختصارات: K، تعداد جمعیت‌های مورد مطالعه. N، تعداد گره‌های شبکه نقشه (محاسبه شده بر اساس داده‌های جمعیت‌های مورد مطالعه)؛ min، حداقل فراوانی روی نقشه؛ max، حداکثر فراوانی روی نقشه؛ avr، میانگین فراوانی روی نقشه؛ std، انحراف استاندارد؛ Rlat، ضریب همبستگی جزئی بین عرض جغرافیایی و فراوانی در نقشه. Rlon، ضریب همبستگی جزئی بین طول و فراوانی بر روی نقشه. Rmul، ضریب همبستگی چندگانه فراوانی با هر دو طول و عرض جغرافیایی بر روی نقشه.