Downloaded from pajoohande.sbmu.ac.ir at 5:31 +0330 on Tuesday September 26th 2017

پژوهنده (مجله پژوهشی دانشگاه علوم پزشکی شهید بهشتی)
سال بیست و یکم، شماره ۴، پی‌درپی ۱۱۲، صفحات ۲۱۰ تا ۲۱۸
مهر و آبان ۱۳۹۵

تاریخ دریافت مقاله: ۹۵/۶/۱۳
تاریخ پذیرش مقاله: ۹۵/۱۰/۱۵

مطالعه گسترده ژنومی در طرح ژنتیک کاردیومتابولیک تهران بستر مناسبی جهت گسترش پزشکی فرادقیق در ایران

دکتر مریم السادات دانشپور[۱]، بهاره صداقتی خیاط[۲]، کامران گیتی[۳]، دکتر فریدون عزیزی[۴]*

۱. دکتری ژنتیک ملکولی؛ مرکز تحقیقات سلولی و مولکولی غدد، پژوهشکده علوم غدد درون‌ریز و متابولیسم دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران
۲. کارشناس ارشد ژنتیک، مرکز تحقیقات سلولی و مولکولی، پژوهشکده علوم غدد درون‌ریز و متابولیسم دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران
۳. کارشناس ارشد ژنتیک، مرکز تحقیقات سلولی و مولکولی، پژوهشکده علوم غدد درون‌ریز و متابولیسم دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران
۴. فوق تخصص غدد درون ریز و متابولیسم، مرکز تحقیقات غدد درون ریز و متابولیسم، پژوهشکده علوم غدد درون‌ریز و متابولیسم دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

چکیده

در چند دهه اخیر، تلاش برای شناسایی نقش تغییرات ژنتیکی و اثر آنان بر پیشگیری و بروز بیماری‌ها یکی از اساسی‌ترین اهداف علم پزشکی شده است. از آنجا که راهنمایی و دستورات پزشکی بر اساس خصوصیات میانگین تنظیم می‌شود که نمی‌تواند برای تعداد زیادی از انسانها مصداق داشته باشد، لذا پزشکی فرادقیق یا پزشکی شخص محور (precision medicine or personal medicin) جهت استفاده از یکسری خصوصیات ژنتیکی، ژنومیکی، اومیک، اپی‌ژنتیکی، بالینی و آزمایشگاهی هر فرد استفاده می‌شود. با پایان پروژه ژنوم برنامه‌ریزی‌های گسترده‌ای در جهت شناخت هر چه بیشتر این کتاب ژنتیکی انجام گردید. نسخه اصلی ژنوم همچون کتابی ویراستاری نشده و غیر قابل خواندن است لذا محققان در سراسر دنیا سعی در استخراج و دسته‌بندی اطلاعات آن دارند. قابل توجه است که هر فرد اطلاعات شخصی خود را در این کتاب به همراه دارد و هیچ دو نسخه‌ی کاملاً مشابهی از آن یافت نمی‌شود. بانک‌های ژنتیکی بزرگی چون HapMap و پروژه ۱۰۰۰ ژنوم به مطالعات گسترده ژنومی کمک می‌نمایند که خوانش این کتاب عظیم را هر چه دقیق‌تر انجام دهند. در ایران نیز بررسی گسترده ژنومی بر روی جمعیت مطالعه قند و لیپید تهران به عنوان بزرگترین و قدیمی‌ترین مطالعه آینده‌نگر انجام گردیده است تا توسط محققان ایرانی گامی کوچک در راستای شناخت ژنوم جمعیت ایرانی برداشته شود. در این مقاله در صدد هستیم که پس از بررسی نقش یافته‌های مطالعات گسترده ژنومی در پزشکی فرادقیق، به معرفی این مطالعه گسترده ژنومی در ایران بپردازیم.

واژگان کلیدی: مطالعه گسترده ژنومی، پزشکی فرادقیق، مطالعه قند و لیپید تهران

لطفاً به این مقاله به صورت زیر استناد نمایید:
Daneshpour MS, Sedaghatikhayat B, Guity K, Azizi F. Genome wide association study in Tehran cardio-metabolic genetic study could promotes precision medicine in Iran. Pejouhandeh 2016;21(4):210-218.

مقدمه

در هنگام مقایسه ژنوم دو فرد، تفاوت‌های بسیاری دیده می‌شود که این تفاوت‌ها حاصل تغییرات ژنومی مانند پلی‌مورفیسم تک نوکلئوتیدی (Single nucleotide variation) و تغییرات بزرگتر مانند حذف (Deletion)، جایگزینی (Insertion) و تغییر در تعداد کپی‌های ژنی (Copy number variation) هستند. هر کدام از این تغییرات می‌توانند منجر به بروز یک فنوتیپ متفاوت در فرد شوند که این فنوتیپ می‌تواند مربوط به صفات ظاهری فرد مانند قد و یا رنگ چشم و یا در ارتباط با خطر بروز یک بیماری باشد. در حدود سال ۲۰۰۰ میلادی، قبل از آنکه مطالعات گسترده ژنومی معرفی شوند، جهت بررسی ارتباط بین ژنوتیپ و فنوتیپ از بررسی توارث (Inheritance) در خانواده و یا مطالعات ژنتیک پیوستگی (Genetic linkage) استفاده می‌شد. این روش‌ها در بررسی بیماری‌هایی که یک ژن عامل بروز آنها بود، بسیار موفق بودند. اما این یافته‌ها در بیماری‌های چند عاملی پیچیده تکرار پذیر نبودند. مطالعات

*نویسنده مسؤول مکاتبات: دکتر فریدون عزیزی؛ فوق تخصص غدد درون‌ریز و متابولیسم؛ مرکز تحقیقات غدد درون‌ریز و متابولیسم، پژوهشکده علوم غدد درون‌ریز و متابولیسم دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران؛ کد پستی: ۱۹۸۵۷۱۷۴۱۳؛ تلفن: ۲۲۴۳۲۵۶۹ (۰۲۱)؛ نمابر: ۲۲۴۱۶۲۶۴ (۰۲۱)؛
پست الکترونیک: azizi@endocrine.ac.ir

Downloaded from pajoohande.sbmu.ac.ir at 5:31 +0330 on Tuesday September 26th 2017

بعدی نشان دادند که آنالیزهای ارتباطی از نوع مورد- شاهدی، قدرت بیشتری در یافتن ژنوتیپ مرتبط دارند، بدین معنا که حضور افزایش و معنا دار یک الل در افراد بیمار نسبت به افراد سالم، می‌تواند نشانه ارتباط حضور آن الل با بیماری باشد. پس از انتشار اولین نسخه از توالی کامل ژنوم انسان، ابزارهای تشخیص ژنتیکی توسعه پیدا کردند و در این راستا بیوبانک‌های متعددی تشکیل گردیدند. انجام پروژه بزرگ نقشه هاپلوتیپی (International HapMap project) منجر به شناسایی پلی‌مورفیسم‌های تک‌نوکلئوتیدی رایج و همچنین بلوک‌های هاپلوتیپی شد و شناخت این نواحی نشان داد که با استفاده از برخی مارکرها می‌توان اطلاعات بیشتری از تغییرات ژنوم به‌دست آورد. با استفاده از این اطلاعات، آرایه‌های ژنوتیپی طراحی گردید و زمینه گسترش مطالعات گسترده ژنومی فراهم شد.

این مطالعات از نوع مورد شاهدی انجام می‌گردد؛ بدین معنا که گروهی از بیماران در مقابل افراد سالم انتخاب می‌شوند و تمامی این افراد برای قسمت عمده پلی‌مورفیسم‌های تک نوکلئوتیدی شناخته شده (تعداد آن بستگی به تکنولوژی مورد استفاده دارد اما به‌طور کلی یک میلیون و یا بیشتر است) تعیین ژنوتیپ می‌شوند، سپس میزان فراوانی اللی هر کدام از این مارکرها در گروه‌های مورد و شاهد مقایسه می‌شود. اساس گزارش میزان اثر (Effect size) بر مبنای محاسبه نسبت شانس (Odd ratio) است بدین معنا که حضور یا عدم حضور یک الل چقدر به حضور یا عدم حضور فنوتیپ مورد بررسی در یک جامعه‌ی مشخص و معلوم، مربوط است. هنگامی که فراوانی اللی در گروه مورد بیشتر از گروه شاهد است، نسبت شانس بیشتر از یک می‌شود. برای گزارش این نسبت از آزمون خی دو (Chi squared) جهت محاسبه مقدار P (P value) استفاده می‌شود. این نوع مطالعه بر اساس به روش‌های دیگری نیز قابل انجام است. یکی از این روش‌ها بررسی فنوتیپ‌های کمی مانند قد، غلظت بیومارکرها و یا بیان ژن است. روش‌های آماری مورد استفاده بسته به میزان نفوذ بیماری و الگوی غالب و یا مغلوبی آن متفاوت است که عموماً با ابزارهای بیوانفورماتیک مانند PLINK (۱) و SNPTEST (۲) انجام می‌شود. همانطور که می‌دانیم افراد مختلف بلوک‌های هاپلوتیپی مشابه دارند و با استفاده از این اطلاعات که در بانک‌های ژنتیکی ذخیره هستند، می‌توان با یکسری از مارکرهای مشخص، توالی این نواحی را تخمین زد و از این رو یکی از مهمترین مراحل در این نوع مطالعات تخمین ژنوتیپ (Imputation of genotypes) در پلیمورفیسم‌هایی است که بررسی نشده‌اند. بدین‌وسیله تعداد مارکرهای مورد بررسی بیشتر شده و به تبع آن قدرت مطالعه افزایش می‌یابد و همچنین امکان انجام آنالیز متا در میان نتایج کوهورت‌های مختلف فراهم می‌شود. نرم افزارهای MACH (۳) و IMPUTE2 (۴) بدین منظور استفاده شده و با روش‌های آماری نتایج اللی حاصل از GWAS را با ژنوم رفرانس مقایسه می‌نمایند.

هدف از ارایه مطالعه مروری حاصل، رسم تصویری شفاف از اهداف و شیوه اجرای مطالعات گسترده ژنومی خواهد بود. همچنین این مقاله به بررسی مزایا و معایب این نوع مطالعات و کاربست آن در پزشکی خواهد پرداخت. بانک‌های اطلاعاتی این نوع مطالعات نیز معرفی خواهد شد و سپس نمونه‌ای از این نوع مطالعات که در یکی از بزرگترین و قدیمی‌ترین کوهورت‌های ایران، مطالعه قند و لیپید تهران، انجام شده است بررسی خواهد شد.

تاریخچه

بررسی ارتباطی گسترده ژنومی اولین بار در سال ۱۹۹۶ توسط Merikangas جهت بررسی ژنتیکی بیماری های چند عاملی پیشنهاد شد (۵) اما چون ایده آن اجرایی نبود، ۱۰ سال طول کشید تا اولین مطالعه از این نوع انجام شود. پروژه تعیین توالی ژنوم انسان (Human Genome Project-HGP) نیز در سال ۱۹۹۰ آغاز شد اما ۱۳ سال به طول انجامید تا نهایتاً در سال ۲۰۰۳، دقیقاً ۵۰ سال بعد از شناسایی اولیه DNA توسط واتسون و گریک (۶)، گزارش کامل توالی ژنوم انسان منتشر شد. این گزارش نشان داد که ژنوم انسان حاوی حدود سه بیلیون جفت باز می‌باشد که این توالی می‌تواند حاوی ۲۰ تا ۲۵ هزار ژن کد کننده‌ی پروتئین باشند. این گزارش (۳) سنگ بنای مهمی را در مطالعه ژنوم انسان بجا گذاشت و به دنبال آن میلیون‌ها پلی‌مورفیسم تک‌نوکلئوتیدی توسط کنسرسیوم SNP (The SNP Consortium- TSC) و کنسرسیوم بین المللی توالی ژنوم انسان (International Human Genome Sequencing Consortium) در بانک‌های داده گزارش گردید (۴).

پلی‌مورفیسم تک نوکلئوتیدی و تغییرات تعداد کپی‌های ژنی

چند شکلی‌ها یا پلی‌مورفیسم‌های تک نوکلئوتیدی، از شایع‌ترین تغییرات ژنتیکی در ژنوم انسان هستند. در حال

Downloaded from pajoohande.sbmu.ac.ir at 5:31 +0330 on Tuesday September 26th 2017

حاضر بالغ بر ۱۰ میلیون SNP در بانک‌های داده ژنتیکی عمومی معرفی شده‌اند که پیش‌بینی می‌شود اغلب آنها تأثیر عملکردی نداشته باشند. این چند شکلی‌های ژنتیکی، اطلاعات ارزشمندی از شاخص‌های ژنتیکی را بیان می‌کنند که می‌توانند با توجه به وقوع آنها در جمعیت (Linkage Disequilibrium) در تشخیص تغییرات منجر به بیماری، مورد استفاده قرار گیرند.

در دست داشتن اطلاعات تعداد زیادی SNP از ژنوم انسانی دید وسیعی از ژنوم را (در مقایسه با سایر مارکرهای مورد استفاده به عنوان شاخص‌های ژنتیکی مثل میکروستلایت‌ها و یا مینی‌ستلایت‌ها) حاصل می‌کند و امکان مطالعه روی تمام ژنوم انسانی را برای محققین میسر می‌سازد. لازم به ذکر است که SNP ها به تنهایی، نه می‌توانند تمامی تغییرات ژنومی انسان را مشخص کنند و نه بیانگر کاندیداهای ژنتیکی برای بیماری‌های پیچیده و مغایرت‌های اثر دارویی هستند.

اخیراً، مطالعات انجام شده روی هزاران مورد تغییر تعداد کپی‌های ژنی (Copy Number Variations- CNVs)، که در ژنوم انسانی قرار دارند، نشانه‌های تازه‌ای از پیچیدگی تغییرات ژنوم انسانی را نمایان کرده است. محققان بر این باورند که CNVs ها نقش مهمی در اساس ژنتیکی بیماری‌های پیچیده برعهده دارند و امید است با اطلاعات به‌دست آمده از SNP در مطالعات GWAS در آینده این مسأله بیشتر مشخص گردد. CNV ها تغییرات ساختاری یا ژنتیکی هستند که با تغییر تعداد کپی قطعات DNA بزرگتر از یک کیلو باز (شامل جهش‌های حذف، الحاق و مضاعف شدن)، این تغییرات را اعمال می‌کنند.

وجود تغییرات تعداد کپی‌های ژنی در ژنوم انسانی اولین بار در سال ۲۰۰۴ گزارش شد (۸،۷). نقشه کلی CNV در سال ۲۰۰۶ انتشار یافت که بیان کننده ۱۵۰۰ ناحیه تغییرات تعداد کپی ژنی بود (۹). تغییرات تعداد کپی ژنی و سایر تغییرات ساختاری مثل جهش واژگونی که از مدتها پیش شناسایی شده بودند، در بانک‌های اطلاعاتی تغییرات ژنومی قرار داده شدند (http://projects.tcag.ca/variation/).

هدف اصلی از تشکیل چنین بانک‌های بین‌المللی داده، ارایه فهرست جامع از ساختار و نوع تغییرات ژنومی است. تأثیر CNV ها بر بیان ژن‌ها نیز گزارش شده است (۱۰). همچنین، نقش این تغییرات در بیماری‌های پیچیده مثل بیماری‌های خود ایمنی، اوتیسم و دو قطبی به اثبات رسیده است (۱۱-۱۴). بحث در مورد CNV ها از حوزه این مقاله خارج است، اما مقالات مروری متعددی در ارتباط با تغییرات ساختاری وجود دارد (۱۵).

پروژه بین المللی HapMap و مفهوم LD

به دنبال اتمام پروژه ژنوم و کشف میلیون‌ها SNP، یکی از این معضلات پیش رو، سختی بررسی تمامی این SNP ها بود. با هدف کاهش تعداد SNP های مورد بررسی و یافتن بلوک‌های هاپلوتیپی، پروژه HapMap در سال ۲۰۰۳ (۱۶) روی ۲۷۰ نمونه ژنومی از جمعیت‌های اروپایی، افریقایی و آسیایی آغاز گردید. فاز اول و دوم در سال‌های ۲۰۰۵ (۱۷) و ۲۰۰۷ تکمیل گردید. نتایج این مطالعه، SNP های نشانه را در سطح ژنوم مشخص نمود به‌طوری که با تعیین ژنوتیپ این SNP ها می‌توان تغییرات دیگر نواحی ژنومی را تخمین زد و این یافته به صرفه‌جویی در وقت و زمان کمک می‌نماید. در نواحی با LD بالا (r^2>0.8)، SNP هایی که در نزدیکی هم قرار دارند، شانس بیشتری دارند که با الگوی توارث به ارث برسند. r^2 بازه‌ای بین صفر تا یک دارد و جهت اندازه‌گیری میزان ارتباط یا LD مابین دو SNP به‌کار می‌رود و بستگی به فراوانی اللی و میزان نوترکیبی بین دو SNP دارد.

پیشرفت‌های کلیدی در تکنولوژی تعیین ژنوتیپ

با توسعه سریع فن‌آوری‌های تعیین ژنوتیپ و کاهش هزینه‌ها، در حال حاضر بسیاری از مؤسسه‌های تحقیقاتی در مدت بسیار کوتاهی می‌توانند بیش از نیم میلیون SNP را تعیین ژنوتیپ نمایند. در این میان می‌توان به چیپ‌های ژنتیکی ساخته شده توسط کمپانی‌های Illumina و Affymetrix اشاره کرد که به دلیل طراحی دقیق آنها قدرت بسیار خوبی برای نقشه برداری از نواحی ژنومی داشته و به‌طور گسترده در مطالعات گسترده ژنومی استفاده می‌شوند. محصولات عرضه شده مانند Illumina HumanHap550 و Affymetrix GeneChip 500K پوشش خوبی از اطلاعات HapMap در دو جمعیت قفقازی و آسیایی ارایه می‌دهند ولی پوشش ژنومی در آفریقا به دلیل تنوع ژنتیکی بیشتر و LD ضعیف‌تر، پایین‌تر است.

یکی دیگر از محصولات جهت ژنوتایپینگ ژنوم Ommini است که توسط شرکت Illumina تولید می‌شود. این محصول در یک چرخه کاری می‌تواند برای یک فرد تا پنج میلیون مارکر را که بر اساس نتایج پروژه ۱۰۰۰ ژنوم به دست آمده‌اند، بررسی کند.

Downloaded from pajoohande.sbmu.ac.ir at 5:31 +0330 on Tuesday September 26th 2017

پروژه ۱۰۰۰ ژنوم

در سال ۲۰۰۷ بعد از بـه اتمـام رسـیدن پـروژه HapMap یـک تـیم بـین‌المللـی بـه جهـت پوشـش نقـص‌هـای پـروژه HapMap تلاش کرد تا کاتالوگی جهان در مورد تفـاوت‌هـای ژنـومی میـان انسـان‌هـا را تولیـد کنـد. از آن‌جـا کـه پـروژه HapMap تنها به بررسی SNP های شایع در جوامع پرداخته بود، پروژه ۱۰۰۰ ژنوم به بررسی گسترده‌تـر و دقیـق‌تـر ژنـوم پرداخت. در این تحقیق، محققان به بررسی ژنوم ۲۵۰۴ نفـر از ۲۶ جامعـه در آفریقـا، شـرق و مرکـز آسـیا، اروپـا و آمریکـا پرداختند. از میان بیش از ۸۸ میلیون منطقه ژنتیکی شناسایی شده، حدود ۱۲ میلیون رایج‌تر بودند که احتمالاً در بسیاری از جوامع مشترک بودند. نتایج این تحقیـق نشـان داد بزرگتـرین تنوع ژنومی در جوامع آفریقایی اسـت کـه بـا شـواهد پیشـین مبنی بر این‌که مهاجرت اولیه بشر از آفریقا آغاز شـده و باعـث ایجاد جوامع دیگر در سراسر جهان شده است، سازگار بود. این کاتالوگ تعداد مناطق شناخته شـده در ژنـوم انسـان را بـه دو برابـر رسـاند و موجـب شـد طیـف وسـیعی از پـژوهش‌هـای زیست‌شناسی و پزشکی انسان مورد استفاده قرار گرفته و پایـه یک درک جدید را از چگونگی مشـارکت تفـاوت‌هـای بـه ارث رسیده در دی-ان-ای با خطر بیماری و واکـنش بـه دارو ارایـه کند.

کنسرسیوم مورد شاهدی Wellcome trust

این مطالعه یکی از بزرگترین مطالعه‌هـای گسـترده ژنـومی است که منجر به کشف مهمترین عوامل ژنتیکی مؤثر بر هفت بیماری شایع در دنیا: اخـتلال دو قطبـی (Bipolar disorder-BD)، بیماری عروق قلبی (Coronary artery disease-CAD)، بیماری کـرون (Crohn's disease-CD)، بیمـاری فشـار خـون (Hypertension-HT)، آرتریـــت روماتوئیـــد (Rheumatoid arthritis-RA)، دیابـت نـوع یـک و دو (Type 1 and 2 diabetes- T1/2D) شـد. در ایـن مطالعـه، هفـده هـزار نفـر بررسی شدند که نمونه‌های انتخاب شده برای هر کدام از هفت بیماری ۲۰۰۰ نفر بود و جهت کنترل، ۳۰۰۰ کنترل مشترک انتخاب گردیدنـد. ایـن تحقیـق یکـی از موفقیـت‌هـای بـزرگ مطالعات گسترده ژنومی به شـمار مـی‌رود چراکـه در طـی آن بسیاری از مناطق جدید ژنومی که با بیماری‌هایی چون دیابـت نوع دو، سرطان و بیماری‌های قلبی در ارتباط بود کشف شدند که در تحقیقات بعدی، اثرگذار بودن آنهـا بـه اثبـات رسـید. از مجموعه لیستی که توسـط مطالعـه WTCCC، معرفـی شـد یکی از جذاب‌ترین نتایج در منطقه ۹p۲۱ بـود کـه ارتبـاط دو ژن CDKN2A/CDKN2B را با بیماری‌های قلبی عروقـی و T2D نشان داد. در همان زمان، ارتباط این منطقـه ژنـومی بـا بیماری عـروق کرونـر قلـب و انفـارکتوس میوکـارد توسـط دو مطالعه مستقل دیگر نیز به اثبات رسید.

کاتالوگ GWAS

مؤسسه ملـی تحقیقـات ژنـوم انسـان (NHGRI) از سـال ۲۰۰۶ شروع به ارایه گزارش‌هایی در مورد ارتباط بین صفات و SNP هـا روی نقشـه ژنـومی انسـان کـرد. ایـن مؤسسـه بـا جمع‌آوری اطلاعات ۱۷۵۱ مطالعه GWAS در سطح جهان و بررسی ۱۱۹۱۲ SNP موفق به انتشار کاتالوگی برای دسترسی عموم شد. در این کاتالوگ ارتباط بین SNP ها و صفات مـورد مطالعه قرار گرفته در کلیه مطالعات GWAS روی نقشه ژنوم مشخص شده‌اند و نام صفاتی که با آن در ارتباط بـوده‌انـد نیـز نوشته شده است. به این طریق می‌توان با اسم صفت مورد نظر به همه اطلاعات ثبت شده ژنتیکی برای آن رسـید (۱۸). ایـن کاتالوگ به طور معمول هر سه ماه به روز می‌شـود و اطلاعـات وارد شده همواره قابل دانلـود کـردن هسـتند. همچنـین ایـن اطلاعات از طریق دیگر دیتابیس‌هـا ماننـد UCSC (۱۹) و یـا Ensembl (۲۱،۲۰) و حتی PheGenI (۲۲) قابـل دسترسـی هستند.

کنترل کیفی داده‌ها

کنترل کیفی داده ها یکی از مهمترین بخش‌های مطالعـات GWAS اسـت چـرا کـه تشـخیص نابجـا و کنـار گذاشـتن نمونه‌های مورد ارزیابی و مارکرهای ژنتیکـی مـی‌توانـد سـبب بـروز انحـراف در نتـایج مطالعـه گـردد. اشـتباه در نتـایج ژنوتایپینگ نیز می‌تواند سبب نمایان شدن ارتباطـات مثبـت و منفی کـاذب و انحـراف سیسـتماتیک نتیجـه مطالعـه گـردد. بسیاری از این ایرادات با افزایش دقت کنترل کیفی روش‌هـای آزمایشگاهی و تعریف نقاط بحرانی نتایج قابل پیشگیری اسـت. انتخاب جامعه مورد بررسی در مطالعات مورد- شاهدی یکی از مراحل مهم قبل از انجام محاسبات آماری اسـت. تفـاوت‌هـای جمعیتی در این نوع مطالعـات بسـیاری مهـم بـوده و بایـد در مراحل قبل از محاسبات، برای آن چاره‌جویی کرد.

اولین مرحله انجام کنترل کیفی از بررسـی نتـایج خروجـی دستگاه‌هـای آزمایشـگاهی بـا بررسـی میـزان و نسـبت دقـت ژنوتیپ آغاز می‌شـود. روش‌هـای سـنتی بـرای بررسـی نتـایج ژنوتایپ در مقیاس‌های کوچک قابل اسـتفاده اسـت امـا بـرای نتایج مطالعات در مقیاس‌های بـالاتر مثـل مطالعـات GWAS

Downloaded from pajoohande.sbmu.ac.ir at 5:31 +0330 on Tuesday September 26th 2017

کاربردی نخواهد بود. الگوریتم بررسی دقت نتایج ژنوتیپ بسته به دستگاه‌های مورد استفاده متفاوت خواهد بود و معمولاً از سوی شرکت‌های سازنده دستگاه‌ها ارایه و با تعریف نقطه برش این دقت محدود خواهد شد. انتخاب نادرست این نقاط می‌تواند سبب حذف اشتباهی قسمت‌های مهم ژنوتایپ شده و باقی ماندن نواحی کم‌اهمیت‌تر (Informative missingness) در بروز یک فنوتیپ گردد. از سوی دیگر انتخاب سخت‌گیرانه این نقاط برش ممکن است سبب کاهش قدرت مطالعه و عمق پوشش ژنوتایپی شود.

در مطالعات GWAS با توجه به حجم بالای مارکرهای اندازه‌گیری شده، نسبت دقت اندازه‌گیری (Call Rate) متفاوت خواهد بود. حذف افراد و مارکرهای دارای نسبت اندازه‌گیری کم بعد از بررسی دقیق نقاط برش، یکی از اولین گام‌های ارزیابی کیفیت داده‌های GWAS است. حذف مارکرها در این گونه مطالعات باید با حداکثر دقت و سخت‌گیری انجام گیرد. در صورت حذف مارکرهای مرتبط با بیماری که در جمعیت کم است، ارتباط آن با بروز بیماری نادیده گرفته خواهد شد که نتیجه آن پایین آمدن قدرت مطالعه است.

بررسی افراد مورد مطالعه بعد از تأیید نتایج آزمایشگاهی ژنوتیپ صورت خواهد گرفت که به‌طور کلی شامل ۴ قسمت است: ۱) تطبیق جنسیت اعلام شده نمونه و نتیجه تعیین جنسیت از روی اطلاعات GWAS، ۲) مشخص کردن افراد دارای نتایج ژنوتایپ با دقت کم، ۳) مشخص کردن دوقلوهای همسان، نمونه‌های تکراری و نمونه‌های اشتباه آزمایشگاهی، ۴) پیدا کردن افراد دارای نیای مشترک.

بررسی مارکرهای مورد مطالعه نیز شامل ۳ قسمت است: ۱) یافتن SNP با درصد بالای ژنوتایپ ناموفق، ۲) پیدا کردن SNP هایی که از قانون هاردی واینبرگ (HWE: Hardy-Weinberg equilibrium) تبعیت نمی‌کنند، ۳) جدا کردن مارکرهای دارای فراوانی اللی بسیار کم.

آنالیز بیوانفورماتیک

بسیاری از برنامه های استاندارد محاسبات آماری (R (۲۳ و SPSS) برای محاسبات اولیه قابل بهره‌برداری است اما در مواجه با داده‌های GWAS با توجه به حجم بالای این نوع داده‌ها، بسیاری از برنامه‌های رایج غیر قابل استفاده خواهند بود. در این شرایط بسیاری از محققین از برنامه‌های متن باز و قابل تغییر استفاده می‌کنند که از جمله می‌توان به برنامه‌های PLINK و GenABEL (۲۴) اشاره کرد. نگهداری از داده‌های حجیم GWAS به طرق مختلف، کاهش حجم دیسک اشغالی توسط برنامه، انجام مراحل پیش آزمون به صورت خودکار، امکان دستکاری داده‌های SNP و امکان انجام بسیاری از محاسبات معمول از جمله دلایل استفاده از این برنامه‌ها است.

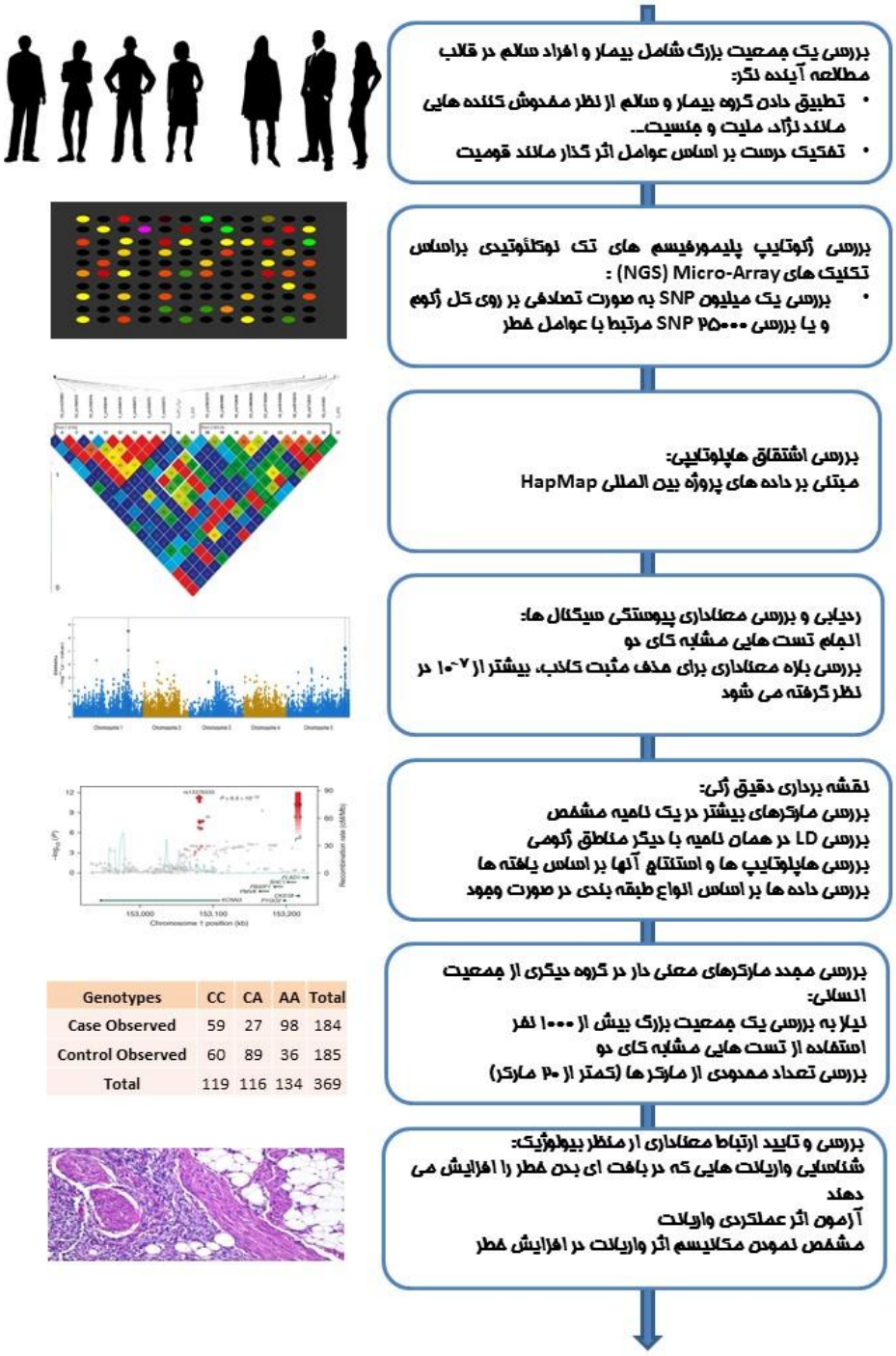
در دسته‌بندی و تعریف زیر مجموعه‌های جمعیت مورد مطالعه، نیاز به داشتن جمعیت مرجع جهت مقایسه و تقسیم‌بندی وجود دارد و معمولاً از جمعیت‌های ارایه شده از مطالعه Hapmap (Hapmap Project Populations) و 1000 Genome (1000 Genome Project) استفاده می‌شود. برای بررسی اطلاعات SNP های مطالعات GWAS بانک‌های داده مختلفی ارایه گردیده است. بانک‌های داده DBSNP (۲۵)، SNPedia (۲۶)، UCSC (۲۷)، Ensemble (۲۸) از این جمله‌اند.

کاربرد در پزشکی فرا دقیق

تا به امروز پزشکی یک درمان ثابت برای تمام بیماران با یک تشخیص معین را پیشنهاد می‌نمود، اما بر اساس یک بیمار در حد میانگین، دستورالعمل درمان بیماری را مشخص می‌کرد اما امروزه مشخص شده است که تعداد کمی از بیماران در حد میانگین هستند. ژن‌های منحصر به فرد، محیط و شیوه‌ی زندگی باعث می‌شوند که حساسیت‌پذیری افراد جهت ایجاد بیماری و پاسخ به درمان متفاوت باشد. امروزه رویکرد علوم پزشکی از سمت درمان به سمت پیشگیری تغییر کرده است، لذا در این راستا چهار محور پزشکی پیشگویانه، پزشکی شخص محور، پزشکی پیشگیرانه و پزشکی مشارکت گسترش یافته‌اند (۲۹). در هر فرد بیمار، شناخت ساختار بیماری، نحوه عمل، فلور طبیعی، مکانیسم تأثیر دارو روند پیشگیری، تشخیص و درمان را در بر می‌گیرد. همان‌طور که می‌دانیم بسیاری از بیماری‌های چند عاملی از طریق داشتن اطلاعات ژنتیکی می‌توانند قابل پیش‌بینی باشند؛ از این‌رو مطالعات گسترده ژنومی می‌توانند در هر جمعیتی پس از یافتن نواحی ژنتیکی مرتبط با بیماری و محاسبه میزان اثر مداخلات محیطی، میزان ریسک فردی در ابتلا به بیماری را محاسبه نمایند و یک درمان پیشگیری‌کننده‌ی مناسب همراه با تغییر شیوه‌ی زندگی پیشنهاد نماید. در همین راستا استفاده از یکسری خصوصیات ژنتیکی، ژنومیکی، اومیک، اپی‌ژنتیک، بالینی و آزمایشگاهی هر فرد در حوضه پزشکی فرادقیق یا پزشکی شخصی (precision medicine or personal medicin) استفاده می‌شود.

Downloaded from pajoohande.sbmu.ac.ir at 5:31 +0330 on Tuesday September 26th 2017

اطلاعات ژنتیکی حاصل از این مطالعات، بستر گسترده‌ای را جهت مدل‌سازی‌های کامپیوتری با هدف کشف ژن‌های جدید در این زمینه فراهم می‌نمایند. اغلب مقاله‌هایی که یافته‌های آنالیزهای گسترده‌ی ژنومی را گزارش می‌کنند، لیستی از پلی‌مورفیسم‌هایی را که به‌طور معنی‌داری با فنوتیپ مورد نظر در ارتباط هستند، گزارش می‌کنند و در پی یافتن مدل‌های ریسک نیستند. لذا با یافتن مدل‌های ریسک از نتایج این مطالعات می‌توان گامی به سوی هموارسازی پزشکی فرادقیق برداشت. یافته‌های ژنومیک از مطالعات GWAS در کنار دیگر عوامل مداخله‌کننده به پزشکی فرادقیق کمک می‌کند تا راهبردهای درمانی ویژه هر فرد مشخص شود. تاکنون اطلاعات مربوط به ۲۱۲۹ مطالعه GWAS انجام شده در سراسر دنیا در کاتالوگ GWAS ثبت گردیده و یافته‌های این مطالعات بستر مناسبی را برای پزشکی فرادقیق در موضوع ژنومیک فراهم کرده است.



شکل ۱. روند انجام مطالعه گسترده ژنومی.

Downloaded from pajoohande.sbmu.ac.ir at 5:31 +0330 on Tuesday September 26th 2017

مطالعه ژنتیک کاردیومتابولیک تهران

همانطور که در این نوشتار ذکر گردید، جهت پیشگیری، تشخیص و درمان بیماری‌های چند عاملی مؤثر بر سلامت جامعه، نیاز به شناسایی عوامل اثرگذار محیطی و ژنتیکی است. با پیشرفت سریع علم ژنتیک و توسعه دستگاه‌ها و نرم‌افزارهای مربوطه، حجم زیادی از داده‌های ژنتیک جمعیتی تولید و در بانک‌های اطلاعاتی ذخیره شده است. نتایج مطالعات گسترده ژنومی نشان داده است که عوامل مستعدکننده بیماری‌های ژنتیک چند عاملی مرتبط با بروز بیماری در اقوام مختلف متفاوت است و یافته‌های هر جمعیت باید در مطالعه مشابه در جمعیت‌های دیگر تأیید گردد.

مطالعه آینده‌نگر قند و لیپید تهران با رویکردی جامعه‌نگر، بر پایه اطلاعات اولیه مبنی بر شیوع پیشرونده بیماری‌های غیرواگیر، روی بیش از ۱۵۰۰۰ نفر از جمعیت ساکن در شرق تهران انجام شده است و تاکنون در ۶ مقطع زمانی با بازه‌های سه ساله از این افراد اطلاعات جمع‌آوری شده است. این مطالعه، قدیمی‌ترین، بزرگترین و معتبرترین مطالعه‌ی آینده‌نگر در ایران است (۳۰). جزئیات طراحی این مطالعه به تفصیل شرح داده شده است (۳۱). جزئیات مطالعه گسترده‌ی ژنومی در مطالعه ژنتیک کاردیومتابولیک تهران که روی بیش از ۱۴۰۰۰ نفر از افراد شرکت‌کننده در طرح قند و لیپید تهران انجام شده است در مجله JMIR Research Protocols چاپ شده است (۳۲). در مطالعه مذکور، کلیه افرادی که تحت پوشش مطالعه قند و لیپید تهران قرار گرفته‌اند، به‌صورت خانوادگی بررسی شدند. در مجموع، ۳۹۵۹ خانواده شامل ۱۳۸۸۷ نفر در مطالعه گسترده ژنومی بررسی و هر نفر برای بیش از ۷۰۰ هزار مارکر ژنتیکی تعیین ژنوتیپ شده است. از میان ۱۹۹۳ فنوتیپ بررسی شده در کلیه GWAS هایی که تاکنون انجام شده‌اند و اطلاعات آن در کاتالوگ GWAS موجود است، ۲۰۹ فنوتیپ آن در این جمعیت بررسی شده و اطلاعات فنوتیپی و ژنوتیپی آن موجود است. امکان تعریف و یا اندازه‌گیری ۴۰۵ فنوتیپ دیگر نیز در نمونه‌های سرمی این افراد وجود دارد.

در طی سال‌های گذشته، متخصصین و صاحب نظران در پژوهشکده علوم غدد درون‌ریز و متابولیسم دانشگاه علوم پزشکی شهید بهشتی در زمینه تأثیر تغییرات رژیم غذایی (۳۹-۳۳)، شیوع دیابت و تغییرات میزان لیپید (۴۲-۴۰)، فنوتیپ‌های مرتبط با چاقی (۴۵-۴۳) و عوامل ژنتیکی اثرگذار بر این تغییرات (۴۷،۴۶،۴۴) تحقیقات زیادی انجام داده‌اند. در همکاری‌های بین‌المللی، مقالات مشترکی روی این جمعیت در مجلات کاملاً معتبر بین‌المللی چاپ شده است (۴۸, ۴۹). با وجود این زیرساخت و افزوده شدن داده‌های حاصل از فناوری امیکس می‌توان گام مؤثری در جهت پزشکی فرادقیق برداشت.

REFERENCES

1. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, *et al*. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 2007; 81 (3): 559-75.

2. Marchini J, Howie B. Comparing algorithms for genotype imputation. Am J Hum Genet 2008; 83(4): 535-9.

3. Finishing the euchromatic sequence of the human genome. Nature 2004; 431(7011): 931-45. Epub 2004/10/22.

4. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, *et al*. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001; 409(6822): 928-33. Epub 2001/03/10.

5. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273(5281): 1516-7.

6. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953; 171(4356): 737-8. Epub 1953/04/25.

7. Sebat J, Lakshmi B, Troge J. Large-scale copy number polymorphism in the human genome. Science 2004; 305:525-8.

8. Iafrate AJ, Feuk L, Rivera MN. Detection of large-scale variation in the human genome. Nat Genet 2004; 36: 949-51.

9. Redon R, Ishikawa S, Fitch KR. Global variation in copy number in the human genome. Nature 2006; 444: 444-54.

10. Stranger BE, Forrest MS, Dunning M. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007; 315: 848-53.

11. Fanciulli M, Norsworthy PJ, Petretto E. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. Nat Genet 2007; 39: 721-3.

12. Yang Y, Chung EK, Wu YL. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE). Am J Hum Genet 2007; 80: 1037-54.

Downloaded from pajoohande.sbmu.ac.ir at 5:31 +0330 on Tuesday September 26th 2017

13. Sebat J, Lakshmi B, Malhotra D. Strong association of de novo copy number mutations with autism. Science 2007; 316: 445-9.

14. Lachman HM, Pedrosa E, Petruolo OA. Increase in GSK3beta gene copy number variation in bipolar disorder. Am J Med Genet B Neuropsychiatr Genet 2007; 144: 259-65.

15. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 2006; 7: 85-97.

16. Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, *et al*. The international HapMap project. Nature 2003; 426(6968): 789-96. Epub 2003/12/20.

17. Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site. Genome research. 2005;15(11):1592-3. Epub 2005/10/28.

18. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, *et al*. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2014; 42(Database issue): D1001-6. Epub 2013/12/10.

19. Karolchik D, Hinrichs AS, Kent WJ. The UCSC Genome Browser. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [*et al*]. 2012; Chapter 1: Unit1 4. Epub 2012/12/21.

20. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, *et al*. Ensembl 2013. Nucleic Acids Res 2013; 41(Database issue): D48-55. Epub 2012/12/04.

21. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, *et al*. Ensembl 2012. Nucleic Acid Res 2012; 40(Database issue): D84-90. Epub 2011/11/17.

22. Ramos EM, Hoffman D, Junkins HA, Maglott D, Phan L, Sherry ST, *et al*. Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. Eur J Hum Genet EJHG 2014; 22(1): 144-7. Epub 2013/05/23.

23. Team RC. R: A language and environment for statistical computing. 2013.

24. Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics 2007; 23(10): 1294-6.

25. Sherry ST, Ward M-H, Kholodov M, Baker J, Phan L, Smigielski EM, *et al*. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001; 29(1): 308-11.

26. Cariaso M, Lennon G. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. Nucleic Acids Res 2012; 40(D1):D1308-D12.

27. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu Y, *et al*. The UCSC genome browser database. Nucleic Acids Res 2003; 31(1): 51-4.

28. Dietterich TG, editor. Ensemble methods in machine learning. International workshop on multiple classifier systems. Springer; 2000.

29. Hood L, Friend SH. Predictive, personalized, preventive, participatory (P4) cancer medicine. Nat Rev Clin Oncol 2011; 8(3): 184-7.

30. Kheradmand M, Enayati A, Rafiei A, Moosazadeh M. Population based cohort studies in Iran: a review article. J Mazandaran Univ Med Sci 2015; 25(125): 171-82.

31. Azizi F, Ghanbarian A, Momenan AA, Hadaegh F, Mirmiran P, Hedayati M, *et al*. Prevention of non-communicable disease in a population in nutrition transition: Tehran Lipid and Glucose Study phase II. Trials. 2009; 10:5. Epub 2009/01/27.

32. Daneshpour M, Fallah M-S, Sedaghati-Khayat B, Guity K, Khalili D, Hedayati M, *et al*. Rationale and design of Genetic study in Cardio-Metabolic risk factors: Tehran Cardio-Metabolic Genetic Study (TCGS). JMIR Res Protoc 2017.

33. Asghari G, Eftekharzadeh A, Hosseinpanah F, Ghareh S, Mirmiran P, Azizi F. Instability of different adolescent metabolic syndrome definitions tracked into early adulthood metabolic syndrome: Tehran Lipid and Glucose Study (TLGS). Pediatric diabetes. 2016. Epub 2016/01/31.

34. Bahadoran Z, Mirmiran P, Tahmasebi Nejad Z, Ghasemi A, Azizi F. Serum nitric oxide is associated with the risk of chronic kidney disease in women: Tehran lipid and glucose study. Scand J Clin Lab Invest 2016; 76(4): 304-8. Epub 2016/03/10.

35. Cheraghi Z, Mirmiran P, Mansournia MA, Moslehi N, Khalili D, Nedjat S. The association between nutritional exposures and metabolic syndrome in the Tehran Lipid and Glucose Study (TLGS): a cohort study. Public health. 2016. Epub 2016/08/09.

36. Farahmand M, Tehrani FR, Amiri P, Azizi F. Barriers to healthy nutrition: perceptions and experiences of Iranian women. BMC Public Health 2012; 12: 1064. Epub 2012/12/12.

37. Mirmiran P, Azadbakht L, Azizi F. Dietary quality-adherence to the dietary guidelines in Tehranian adolescents: Tehran Lipid and Glucose Study. Int J Vitamin Nutr Res. Internationale Zeitschrift fur Vitamin- und Ernahrungsforschung. Journal international de vitaminologie et de nutrition. 2005; 75(3): 195-200. Epub 2005/07/21.

38. Mirmiran P, Bahadoran Z, Golzarand M, Asghari G, Azizi F. Consumption of nitrate containing vegetables and the risk of chronic kidney disease: Tehran Lipid and Glucose Study. Ren Fail 2016; 38(6): 937-44. Epub 2016/04/09.

39. Mirmiran P, Ejtahed HS, Bahadoran Z, Bastan S, Azizi F. Sugar-sweetened beverage consumption and risk of general and abdominal obesity in Iranian adults: Tehran Lipid and Glucose Study. Iran J Public Health 2015; 44(11): 1535-43. Epub 2016/01/09.

40. Bandarian F, Daneshpour MS, Hedayati M, Naseri M, Azizi F. Identification of sequence variation in the apolipoprotein A2 gene and their relationship with serum high-density lipoprotein cholesterol levels. Iran Biomed J 2016; 20(2): 84-90. Epub 2015/11/22.

41. Moslehi N, Shab-Bidar S, Mirmiran P, Sadeghi M, Azizi F. Associations between dairy products consumption and risk of type 2 diabetes: Tehran lipid and glucose study. Int J Food Sci Nnutr 2015; 66(6): 692-9. Epub 2015/05/07.

42. Mozaffary A, Asgari S, Tohidi M, Kazempour-Ardebili S, Azizi F, Hadaegh F. Change in fasting plasma glucose and incident type 2 diabetes mellitus: results from a prospective cohort study. BMJ open 2016; 6(5): e010889. Epub 2016/05/25.

43. Akbarzadeh M, Moghimbeigi A. Trajectories of change in obesity among tehranian families: multilevel latent growth curve modeling. Int J Fam Med 2016; 2016: 2639624.

44. Faam B, Zarkesh M, Daneshpour MS, Azizi F, Hedayati M. The association between inflammatory markers and obesity-related factors in Tehranian adults: Tehran lipid and glucose study. Iran J Basic Med Sci 2014; 17(8): 577-82. Epub 2015/02/26.

45. Mahjub H, Soltanian AR, Daneshpour M, Morris N, Barzin M, Keihani S, *et al*. Rising trends of obesity and abdominal obesity in 10 years of follow-up among Tehranian adults: Tehran Lipid and Glucose Study (TLGS). Int J Fam Med 2015; 18(16): 2981-9. Epub 2016/04/05.

46. Fallah MS, Sedaghatikhayat B, Guity K, Akbari F, Azizi F, Daneshpour MS. The relation between metabolic syndrome risk factors and genetic variations of apolipoprotein V in relation with serum triglyceride and HDL-C level. Arch Iran Med 2016; 19(1): 46-50. Epub 2015/12/26.

47. Zarkesh M, Daneshpour MS, Faam B, Fallah MS, Hosseinzadeh N, Guity K, *et al*. Heritability of the metabolic syndrome and its components in the Tehran Lipid and Glucose Study (TLGS). Genet Res 2012; 94(6): 331-7. Epub 2013/02/05.

48. Holm H, Stefansson K, Gretarsdottir S, Helgason H, Helgadottir A, Sigurdsson A, *et al*. A splice region variant in LDLR lowers non-high density lipoprotein cholesterol and protects against coronary artery disease. Nat Genet 2015; 11(9): e1005379. Epub 2016/05/03.

49. Steinthorsdottir V, Thorleifsson G, Sulem P. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat Genet 2014; 46(3): 294-8. Epub 2015/09/04.

Downloaded from pajoohande.sbmu.ac.ir at 5:31 +0330 on Tuesday September 26th 2017