Archive of SID

**Iranian Journal of Animal Science**
Vol 49, No 1, Spring 2018 (73-81)
DOI: 10.22059/ijas.2018.227090.653506

**علوم دامی ایران**
دورهٔ ۴۹، شمارهٔ ۱، بهار ۱۳۹۷ (ص ۸۱-۷۳)

# تأثیر اسناد کردن و تراکم نشانگری در بهبود درستی ارزیابی‌های ژنگانی

**ناهید پرنا[۱]، اردشیر نجاتی جوارمی[۲]، مصطفی صادقی[۲*] و رستم عبداللهی آرپناهی[۳]**
**۱ و ۲. دانشجوی کارشناسی ارشد و دانشیار، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، کرج**
**۳. استادیار، پردیس ابوریحان دانشگاه تهران**
**(تاریخ دریافت: ۱۳۹٦/٥/۱٤ - تاریخ پذیرش: ۱۳۹۷/۲/۱۱)**

## چکیده

**با دسترسی به اطلاعات نشانگرهای تک نوکلئوتیدی (SNP)، پیش‌بینی شایستگی ژنتیکی افراد بدون اطلاعات پدیدگانی (فنوتیپی) با استفاده از ارزیابی ژنگانی (ژنومی) امکان‌پذیر شده است. لیکن، استفاده از تراشه (چیپ)های متراکم برای ارزیابی همهٔ افراد جمعیت مقرون‌به‌صرفه نیست. برای دستیابی به بیشترین بازدهی در بررسی‌های ژنگانی می‌توان گروهی از حیوان‌ها را با تراشه‌های با تراکم بالا و دیگر افراد را با تراشه‌های کم‌تراکم تعیین نژادگان (ژنوتیپ) کرد، سپس آن‌ها را به تراکم بالا اسناد (ایمپیوت) کرد. در این بررسی، تأثیر سه تراشهٔ کم‌تراکم (1k، 2k و 4k)، اسنادشده به تراکم بالای 10k و رابطهٔ خویشاوندی بین جمعیت مرجع و جمعیت‌های تأیید بر درستی ارزیابی ژنگانی و نیز همبستگی ارزش‌های اصلاحی برآوردشده در تراشه‌های مختلف با استفاده از دادهٔ همانندسازی‌شده، بررسی شد. برای ساختن تراشه‌های کم‌تراکم، ۱۰، ۲۰ و ٤۰ درصد از نشانگرهای 10k به‌صورت تصادفی انتخاب شدند. اسناد کردن (ایمپیوتیشن) با استفاده از نرم‌افزار FImpute صورت گرفت. به‌طورکلی با افزایش تراکم نشانگری، همبستگی بین ارزش‌های اصلاحی برآوردشده به‌دست‌آمده از تراشه‌های مختلف افزایش پیدا کرد. به‌گونه‌ای که، درستی ارزیابی ژنگانی تراشهٔ کم‌تراکم 4k و تراشهٔ متراکم 10k همسان بودند. همچنین پس از اسناد کردن نژادگان‌های تراشهٔ 4k به 10k، تفاوتی بین درستی ارزیابی‌های ژنگانی ایجاد نشد. با افزایش فاصلهٔ جمعیت مرجع و جمعیت‌های تأیید، درستی اسناد کردن کاهش یافت.**

**واژه‌های کلیدی:** رابطهٔ خویشاوندی، نشانگر، همانندسازی.

# Impact of genotype imputation and density of markers on the accuracy of genomic prediction

**Nahid Parna[1], Ardeshir Nejati Javaremi[2], Mostafa Sadeghi[2*] and Rostam Abdollahi ArPanahi[3]**
1, 2. M. Sc. Student and Associate Professor, University College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran
3. Assistant Professor, Aboureyhan Campus, University of Tehran, Iran
(Received: Aug. 5, 2017 - Accepted: May 1, 2018)

## ABSTRACT

Using SNP markers information and genomic evaluation approach, predicting the genetic merit of individuals without phenotypic records is now possible. However, using high-density panels for genomic evaluation of all individuals is not economically feasible. To achieve high genomic prediction accuracy with reasonable price, it is possible to genotype a proportion of animals with high-density panels and the rest of animals with low-density panels then impute them to high-density genotypes. In this study, the effect of three low-density panels (1k, 2k and 4k), genotype imputation to 10k panel and the relationship between reference and validation populations on the accuracy of genomic predictions and also the correlation between the estimated breeding values using panels with different densities in simulated data were assessed. The low density panels genotypes were actually consisting of 10, 20 and 40 percent of 10k markers selected randomly and FImpute was used for genotype imputation. As a general trend, by increasing the density of markers, the correlation between the estimated breeding values was increased using different panels. So, the accuracy of genomic predictions was similar using 4k and 10k genotypes. Moreover, imputing 4k to 10k genotypes, did not improve the accuracy of genomic prediction. However, the accuracy of estimated breeding values was increased after imputation from 1k or 2k to 10k. The accuracy of imputation was decreased when the reference and validation populations were more distant.

**Keywords:** Genetic relationship, marker, simulation.

* Corresponding author E-mail: sadeghimos@ut.ac.ir

www.SID.ir

Archive of SID



## مقدمه

با پیشرفت در فناوری‌های ژنگانی (ژنومی) و استفاده از انتخاب ژنگانی، پس از برآورد تأثیر نشانگرها می‌توان بدون استفاده از اطلاعات پدیدگانی (فنوتیپی)، شایستگی ژنتیکی افراد را پیش‌بینی کرد. انتخاب ژنگانی موجب افزایش پیشرفت ژنتیکی به‌واسطهٔ کاهش فاصلهٔ نسل به‌وسیلهٔ امکان انتخاب اولیه و زودهنگام حیوانات بر پایهٔ ارزش اصلاحی ژنگانی را فراهم می‌کند (Schaefer, 2006). در ارزیابی‌های ژنگانی با استفاده از رابطهٔ بین پدیدگان با شمار بسیار زیادی نشانگر تک نوکلئوتیدی[1] (SNP) که در کل ژنگان پراکنده است، ارزش اصلاحی برآوردشدهٔ ژنگانی (GEBV)[2] حیوان‌ها پیش‌بینی می‌شود (Meuwissen *et al*., 2001). به‌طورکلی ارزیابی ژنگانی به سه مرحله تقسیم می‌شود (Goddard & Hayes, 2007): ۱- تعیین نژادگان (ژنوتیپ) حیوان‌ها بر پایهٔ تراشهٔ SNP، ۲- برآورد اثر SNPها به‌طور همزمان برای صفت مورد بررسی با استفاده از روش‌های مختلف آماری در جمعیت مرجع[3]، ۳- محاسبهٔ ارزش برآوردشدهٔ اصلاحی ژنگانی برآورد شده در جمعیت تأیید[4]. جمعیت مرجع، جمعیتی است که افراد آن نژادگان شده و رکورد پدیدگانی برای صفت مورد بررسی دارند. جمعیت تأیید، جمعیتی است که نژادگان افراد تعیین‌شده ولی بدون رکورد پدیدگانی برای صفت مورد بررسی هستند. در بررسی‌های انتخاب ژنگانی و پویش ژنگان (GWAS)[5] به شمار زیاد افراد نژادگان‌شده با تراکم نشانگری بالا نیاز است. یکی از عامل‌های مؤثر بر درستی پیش‌بینی‌های ژنگانی و GWAS، تراکم نشانگر است. به‌طور عمده، تراکم بالاتر منجر به پیش‌بینی بهتر و نقشه‌یابی دقیق‌تر QTL می‌شود که علت آن نبود تعادل پیوستگی (LD)[6] قوی‌تر بین نشانگر و جهش علی است. چنانچه بتوان هزینهٔ تعیین نژادگان را کاهش داد می‌توان برنامهٔ تعیین نژادگان همهٔ افراد گله‌های تجاری را در برنامه‌های مدیریتی مزارع به‌صورت منظم به‌کار برد. بدین ترتیب تصمیم‌گیری مدیریتی در حذف و جایگزینی در گله‌های گاو شیری بر مبنای تعیین نژادگان تلیسه‌ها با تراشه (چیپ)های تراکم پایین و سپس اسناد کردن (ایمپیوت) نژادگان‌های گمشده از افراد جمعیت مرجع خواهد بود. یک راهکار برای بهبود پیش‌بینی‌های ژنگانی حیوانات استفاده از تراشه‌های با تراکم پایین آنگاه اسناد کردن آن به تراشه‌های متراکم‌تر است (Berry *et al*., 2013).

گروهی از تراشه‌های SNP تجاری برای گاو شیری موجود است. این تراشه‌ها شامل BovineLD با تراکم ۶۹۰۹ نشانگر، BovineLD با تراکم نشانگری ۸۶۵۵، BovineSNP50v2 با تراکم ۵۴۶۰۹ نشانگر و BovineHD با تراکم ۷۷۷۹۶۲ نشانگر توسط Illumina تولیدشده است (Wiggans *et al*., 2011). کوچک بودن اندازهٔ جمعیت مؤثر و ساختار جمعیت در گونه‌های دامی امکان اسناد کردن از تراشه با تراکم پایین به تراشه با تراکم بالا را با درستی بالا و بدون نیاز به جمعیت مرجع بسیار بزرگ برای اسناد کردن، فراهم کرده است.با توجه به اینکه اسناد کردن نژادگانی اهمیت زیادی در بررسی‌های ژنگانی گاوهای شیری دارد، نوشتار پیشرو برای پاسخ‌گویی به این پرسش که چه نسبت و تراکمی از نشانگرها می‌تواند برای ارزیابی‌های ژنگانی افزون بر کاهش هزینه‌ها استفاده شود، تدوین شده است.

در این بررسی در آغاز به کمک همانندسازی، داده‌هایی با قابلیت استفاده در زمینهٔ بررسی اسناد کردن نژادگانی ایجاد شد. آنگاه به کمک اطلاعات همانندسازی‌شده، تراشه‌های کم تراکم مختلفی برای اسناد کردن به تراشهٔ متراکم 10k ساخته و عامل‌هایی مانند؛ درستی اسناد کردن هر یک از تراشه‌های استفاده‌شده، درستی ارزیابی‌های ژنگانی تراشه‌های با تراکم مختلف در فرایند افزایش فاصلهٔ جمعیت مرجع با هر یک از جمعیت‌های تأیید، درستی ارزش‌های اصلاحی برآوردشدهٔ به‌دست‌آمده از هر یک از تراشه‌های استفاده‌شده و تأثیر رابطهٔ ژنتیکی بین جمعیت مرجع و جمعیت‌های تأیید بررسی شدند.

1. Single nucleotide polymorphism
2. Genomic estimated breeding value
3. Training population
4. Validation population
5. Genome-wide association study
6. Linkage disequilibrium

www.SID.ir

Archive of SID



## مواد و روش‌ها

**هماننداسازی جمعیت پایه**

برای انجام این بررسی، ژنگانی متشکل از پنج جفت کروموزوم، هرکدام به طول ۱۰۰ سانتی مورگان با شمار ۲۱۰۰ نشانگر SNP با فاصله‌های یکسان از یکدیگر در طول ژنگان و با ۱۰۰ جایگاه ژنی یا QTL روی هر کروموزوم که تنها اثر ژنی افزایشی و توزیع نرمال داشتند، به‌صورت یکنواخت پراکنده شد. نشانگرها و QTLها به‌صورت دو آللی با فراوانی اولیه ۰/۵ بودند. در این بررسی جمعیت پایه با اندازهٔ مؤثر ۲۰۰ فرد (۱۰۰ نر و ۱۰۰ ماده) در نظر گرفته شد. به‌منظور ایجاد نبود تعادل پیوستگی بین نشانگر و QTL، بین افراد نسل پایه ۱۰۰ نسل آمیزش تصادفی انجام شد. برای هماننداسازی نمونه‌گیری مندلی[1] در هر نسل، نوترکیبی بین هر دو رشتهٔ کروموزومی هر والد رخ داد که منجر به تولید نمونه‌های تصادفی تک‌جور (هاپلوتیپ) پدری و مادری برای نتاج نسل بعد شد. از رخداد جهش در این جمعیت صرف‌نظر شد. در این ۱۰۰ نسل اندازهٔ مؤثر جمعیت هر نسل به شکل ثابت و معادل با ۲۰۰ فرد باقی ماند و در این حالت رانش و نوترکیبی باعث شدند که میانگین آمارهٔ نبود تعادل پیوستگی ($r^2$) در نسل ۱۰۰ به ۰/۱۸ برسد. پس از ۱۰۰ نسل آمیزش تصادفی، در نسل ۱۰۱ هر نر به‌صورت تصادفی با ۱۰ ماده آمیزش داده شد که منجر به تولید ۱۰۰۰ فرد در هر نسل شد و این ساختار تا نسل ۱۰۶ ادامه یافت. افراد نسل ۱۰۱ به‌عنوان جمعیت مرجع که هر دو رکورد پدیدگانی و نژادگانی را داشتند و افراد نسل‌های ۱۰۲ تا ۱۰۶ نیز به‌عنوان جمعیت تأیید در نظر گرفته شدند که تنها اطلاعات نژادگانی داشتند.در هر نسل از جمعیت‌های تأیید، ۲۰ درصد افراد نر برتر با استفاده از ارزش‌های اصلاحی برآورد شده با روش بهترین پیش‌بینی ناااریب خطی ژنگانی (GBLUP) و ۸۰ درصد ماده‌ها به‌صورت تصادفی، به‌عنوان والدین نسل بعد انتخاب شدند. وراثت‌پذیری ۰/۵ بود. وراثت‌پذیری، شدّت انتخاب، میزان نوترکیبی و اندازهٔ مؤثر جمعیت در همهٔ پیش‌فرض (سناریو)های هماننداسازی یکسان بود. از بستهٔ نرم‌افزاری hypred (Technow, 2011) برای هماننداسازی جمعیت‌ها استفاده شد.

**مدل آماری**

از روش آماری بهترین پیش‌بینی ناااریب خطی ژنگانی یا GBLUP که توسط VanRaden (2008) پیشنهادشده برای برآورد ارزش‌های اصلاحی در گروه مرجع و آنگاه محاسبهٔ اثر نشانگرها به‌منظور برآورد ارزش‌های اصلاحی در گروه تأیید استفاده شد. مدل آماری استفاده‌شده در روش GBLUP به‌صورت رابطهٔ ۱ است:

$$\mathbf{y} = \mathbf{l}\,\mu + \mathbf{Zu} + \mathbf{e} \qquad (۱)$$

در رابطهٔ بالا $\mathbf{y}$ بردار پدیدگان تصحیح شده برای اثر جنس، $\mu$ میانگین کل، u بردار دربرگیرندهٔ ارزش‌های اصلاحی ژنگانی افراد گروه مرجع است، Z ماتریس ضریب‌های ارتباط‌دهندهٔ مشاهده‌ها به بردار ارزش‌های اصلاحی ژنگانی، e بردار اثر باقیمانده و 1 نیز یک بردار واحد است. عنصرهای ماتریس نژادگانی (X) شامل اعداد ۰، ۱ و ۲ است که نشان‌دهندهٔ شمار رونوشت آلل‌های مرجع هر یک از مکان‌های نشانگری با توجه به نژادگان هر فرد است. در این مدل اثر نشانگرها به‌صورت تصادفی با واریانس یکسان در نظر گرفته شد. رابطه‌های مختلف هندرسون برای برآورد ارزش‌های اصلاحی ژنگانی افراد گروه مرجع ($\widehat{\boldsymbol{u}}$) به‌صورت رابطهٔ ۲ تشکیل شد:

$$\hat{\mathbf{u}} = [\acute{\mathbf{X}}\mathbf{X} + \mathbf{G}^{-1}\mathrm{k}]^{-1}[\acute{\mathbf{X}}\mathbf{y}] \qquad (۲)$$

در رابطهٔ ۲، k برابر است با:

$$\mathrm{k} = \left(\frac{\mathbf{1}-\mathbf{h}^{\mathbf{2}}}{\mathbf{h}^{\mathbf{2}}}\right) \qquad (۳)$$

در این رابطه، $h^2$ وراثت‌پذیری صفت است و هرکدام از عنصرهای ماتریس $\mathbf{G}$ نیز با استفاده از رابطهٔ ۴ محاسبه شد:

$$\mathbf{G}_{\mathrm{jj}} = \frac{1}{\mathrm{m}}\sum_{\mathrm{i}=1}^{\mathrm{m}}\frac{(\mathrm{X}_{\mathrm{ij}}-2\mathrm{p}_{\mathrm{i}})^2}{2\mathrm{p}_{\mathrm{i}}(1-\mathrm{p}_{\mathrm{i}})} \qquad (۴)$$

در این رابطه، $X_{ij}$ شمار کپی آلل مرجع برای iامین SNP در jامین فرد، $p_i$ نیز فراوانی آلل مرجع و m شمار نشانگر است.

به‌منظور برآورد ارزش‌های اصلاحی در گروه تأیید اثر SNP ($\widehat{\boldsymbol{b}}$) با استفاده از ارزش‌های ژنگانی برآورد

1. Mendelian sampling

www.SID.ir

Archive of SID



شده برای افراد گروه مرجع به کمک رابطهٔ ۵ محاسبه شدند.

$$\hat{\boldsymbol{b}} = \mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}\hat{\boldsymbol{u}} \qquad (۵)$$

**اسناد کردن**

روی هر پنج کروموزوم همانندسازی‌شده ۲۱۰۰ SNP قرار گرفت و آنگاه پس از حذف SNPهایی که MAF کمتر از ۵ درصد داشتند، حدود ۱۰۰۰۰ SNP باقی ماند. سپس با انتخاب ۱۰، ۲۰ و ۴۰ درصد از SNPها، به ترتیب تراشه‌های 1k، 2k و 4k از تراشهٔ متراکم 10kایجاد و استفاده شد. نتایج همهٔ پیش‌فرض‌های بالا با پیش‌فرض‌های دارای تراشهٔ متراکم 10k بین افراد مقایسه شدند. لازم به یادآوری است که نشانگرهای ازدست‌رفته با تراکم پایین در جمعیت تأیید موجود بوده و جمعیت مرجع دارای نشانگر ازدست‌رفته در حین تعیین نژادگان با اسنیپ تراشهٔ متراکم نیست. اسناد کردن برای هر پیش‌فرض ده بار تکرار شد و در پایان از نتایج آن‌ها، میانگین‌گیری به عمل آمد.

به‌منظور بررسی تأثیر رابطهٔ بین افراد جمعیت مرجع و جمعیت‌های آزمون بر درستی اسناد کردن، آمیزش به مدت پنج نسل ادامه پیدا کرد. نژادگان حیوانات در هر نسل از جمعیت‌های تأیید برای مقایسهٔ درستی اسناد کردن، به‌وسیلهٔ سه تراشه کم تراکم 1k، 2k و 4k طراحی و با استفاده از نژادگان افراد جمعیت مرجع اسناد شدند. اسناد کردن نژادگانی با استفاده از نرم‌افزار FImpute و مدل افراد خویشاوند انجام شد. نژادگان کل کروموزوم‌ها باهم اسناد شد. برای هر یک از تراکم‌ها و در هر پنج نسل جمعیت تأیید فایل‌های ورودی مورد نیاز برنامهٔ FImpute ساخته شدند (Sargolzaei *et al*., 2014). این نرم‌افزار برای هر بار انجام اسناد کردن به سه فایل ورودی نیاز دارد که اولی شامل اطلاعات نژادگانی افراد، دومی شامل نام و جایگاه کروموزومی SNP روی تراشهٔ متراکم و تراشه با تراکم پایین و سومی شامل اطلاعات شجره‌ای افراد است.

**پیش‌بینی ارزش اصلاحی**

از پدیدگان تصحیح شده برای اثر ثابت جنس، برای برآورد اثر نشانگری در جمعیت مرجع استفاده شد و آنگاه از اثر نشانگری برآورد شده در هر یک از تراکم‌های استفاده‌شده در جمعیت مرجع برای پیش‌بینی ارزش‌های اصلاحی ژنگانی افراد جمعیت تأیید استفاده شد. برای ارزیابی عملکرد تراشه‌های گوناگون و تراکم‌های مختلف بر درستی ارزیابی ژنگانی، پس از حل رابطه‌های بالا و برآورد اثر هر یک از نشانگرها ارزش اصلاحی ژنگانی افراد گروه تأیید از جمع اثر نشانگرها و با توجه به نژادگان آن‌ها به‌صورت رابطهٔ ۶ محاسبه شد:

$$\mathrm{GEBV_i} = \sum_i^m \mathbf{X_{ij}}\,\hat{\mathrm{b}}_j \qquad (۶)$$

درستی برآورد ارزش‌های اصلاحی ژنگانی از همبستگی بین ارزش‌های اصلاحی واقعی و ارزش‌های اصلاحی ژنگانی برآورد شده به دست آمد. تغییرپذیری درستی GEBV برای نسل مرجع و پنج نسل گروه تأیید بررسی شد. هر پیش‌فرض همانندسازی که در برگیرندهٔ تراکم‌های مختلف نشانگرهای SNP و داده‌های ازدست‌رفته بود ده بار تکرار شد و میانگین تکرارها همراه با خطای استاندارد میانگین‌ها ذخیره شد.

**بررسی درستی اسناد کردن**

در این طرح از روش CR (Allele Correct Rate)، برای بررسی درستی روش‌های مختلف اسناد کردن استفاده شد. میزان درستی اسناد کردن برای کل SNPها، از درصد آلل‌هایی که به‌طور درست اسنادشده بودند (بر پایهٔ انطباق[1] نژادگان اسنادشده با نژادگان واقعی برای کل SNPها) به‌کل SNPهای اسنادشده × ۱۰۰ محاسبه شد. این روش به معنای میزان درستی آللی است و چگونگی محاسبهٔ آن بدین‌صورت است:

$$\mathrm{CR} = \left(1 - \frac{\text{Number of errors}}{\text{Total number of masked SNPs genotypes} * 2}\right) * 100 \qquad (۷)$$

## نتایج و بحث

**مقایسهٔ تراشه‌های 1k، 2k و 4k**

درستی ارزیابی ژنگانی در جمعیت مرجع و هر نسل از جمعیت تأیید با توجه به تراکم‌های مختلف محاسبه شد (شکل ۱). نتایج نشان داد، درستی ارزیابی ژنگانی

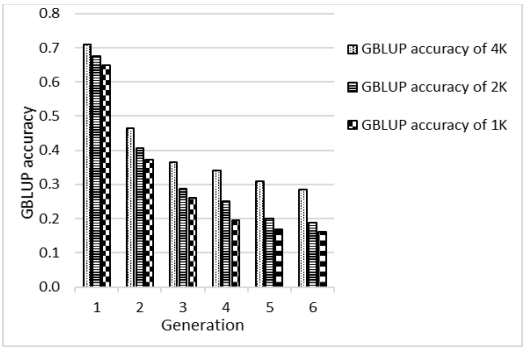
[1]Concordance

www.SID.ir

Archive of SID



هنگامی‌که از تراشهٔ 4k استفاده شود، بیشتر از هنگامی است که از تراشه‌های 2k و 1k استفاده شود. با افزایش فاصلهٔ بین جمعیت مرجع و هر نسل از جمعیت تأیید، درستی ارزش‌های اصلاحی ژنگانی به‌شدت، کاهش می‌یابد. این روند قابل انتظار است زیرا با افزایش فاصلهٔ بین جمعیت مرجع و جمعیت تأیید، وضعیت LD بین QTLها و نشانگرها به‌هم‌خورده و اثر QTLها به‌خوبی نمی‌توانند به‌وسیلهٔ نشانگرها به دام بیفتد. این نتایج با یافته‌های پیشین نیز همخوانی دارد (Habier *et al*, 2007; Ogawa *et al*., 2016; Judge *et al*., 2016). Meuwissen *et al.* (2001) در نتایج بررسی‌های خود نشان دادند، هرچه فاصلهٔ نسلی بین جمعیت مرجع و جمعیت تأیید کمتر باشد درستی ارزش اصلاحی ژنگانی برآورد شده بیشتر خواهد بود. آنان در نتایج بررسی‌های خود بیان کردند، به دلیل برآورد نکردن دوبارهٔ اثر SNP در هر نسل و تغییر میزان و مرحلهٔ (فاز) نشانگر و QTLبه دلیل رخداد نوترکیبی، باگذشت زمان درستی برآوردها کاهش می‌یابد.

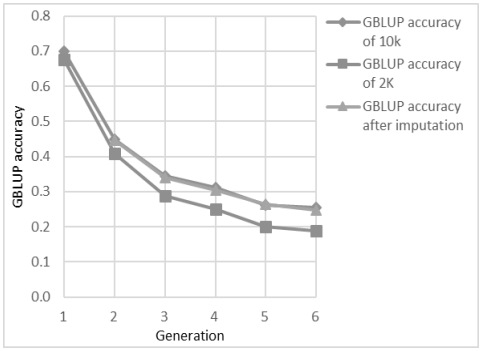



شکل ۱. درستی ارزیابی ژنگانی در تراشه‌های 1k، 2k و 4k در جمعیت مرجع و هر نسل از جمعیت تأیید، پیش از اسناد کردن

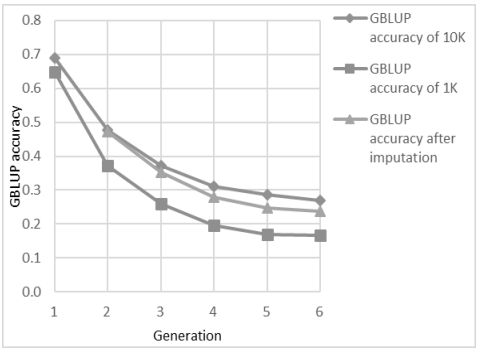
Figure 1. Accuracy of genomic prediction in 1k, 2k and 4k panels in the reference population and each generation of the validation population, before imputation

درستی ارزیابی ژنگانی برای تراشهٔ 1k کمتر از دو تراشهٔ دیگر بود. با این حال تفاوت درستی ارزیابی ژنگانی با تراشهٔ 2k و 4k به‌ویژه در نسل‌های آغازین کم بود و هر چه فاصلهٔ جمعیت تأیید از جمعیت مرجع بیشتر بود تفاوت درستی ارزیابی‌های ژنگانی بیشتر شد. لازم به یادآوری است با افزایش تراکم تراشه‌های کم تراکم، تفاوت بین درستی ارزیابی ارزش‌های اصلاحی برآورد شده با استفاده از تراشهٔ 10k و تراشه‌های 1k، 2k و 4k به ترتیب کاهش یافت (شکل ۲).

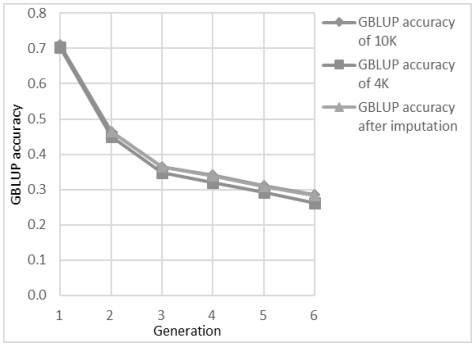
a)




b)




c)




شکل ۲. درستی ارزیابی‌های ژنگانی با استفاده از تراشهٔ 10k، تراشه‌های 1k، 2k و 4k و پس از اسناد تراشه‌های 1k، 2k و 4k به تراشهٔ 10k در شش نسل. نسل اول جمعیت مرجع و نسل دو تا شش جمعیت‌های تأیید یا آزمون هستند.

Figure 2. Accuracy of genomic predictions by using 10k panel, 1k, 2k and 4k panels and after imputing 1k, 2k and 4k panels to 10k panel during six generations. First generation is the reference population and generation two to six are validation or test populations.

www.SID.ir

Archive of SID



یکی از عامل‌های مؤثر بر درستی ارزیابی ژنگانی تراکم نشانگر است. به‌طورکلی، انتظار می‌رود که افزایش تراکم نشانگر منجر به شناخت بهتر مرحلهٔ نبود تعادل پیوستگی بین نشانگر و QTL شده که درنهایت سبب افزایش درستی ارزیابی ژنگانی می‌شود (Solberg *et al*., 2008). VanRaden *et al.* (2009) در نتایج بررسی‌های خود گزارش کردند، افزایش تراکم نشانگر از ۱۰ هزار به ۲۰ هزار و از ۲۰ هزار به ۴۰ هزار سبب افزایش درستی برآورد ارزش اصلاحی برای همهٔ صفات شده است. درحالی‌که گزارش شده است با افزایش تراکم SNP به بیشتر از ۵۰ هزار، درستی ارزیابی ژنگانی به میزان ناچیزی افزایش می‌یابد (Harris *et al*., 2011). ازآنجاکه هدف از افزایش تراکم، افزایش نبود تعادل پیوستگی بین نشانگر و QTL است، به نظر می‌رسد که الگوی نبود تعادل پیوستگی نقش کلیدی در این زمینه داشته باشد (Bolormaa *et al*., 2015).

**بررسی درستی اسناد کردن**

میانگین نرخ درستی اسناد کردن نژادگان‌ها برای اسناد نژادگان‌های تراشه‌های 1k، 2k و 4k از نژادگان‌های تراشهٔ 10k بر پایهٔ میانگین ده تکرار و هر تکرار شامل پنج کروموزوم، در پنج نسل در جمعیت‌های تأیید یا آزمون در شکل ۳ نشان داده شده است. روند کاهش درستی اسناد کردن برای تراشهٔ 4k آهسته‌تر از تراشه‌های 2k و 1k بود. ازآنجاکه در استفاده از تراشه‌های تجاری، نژادگان واقعی افراد مشخص نیست از معیار درستی اسناد کردن برای ارزیابی نژادگان‌های پیش‌بینی‌شده استفاده می‌شود. درستی اسناد کردن بدون دانستن نژادگان واقعی و با استفاده از احتمال‌های نژادگان پسین برای نژادگان‌های اسنادشده برآورد می‌شود. در همهٔ تراشه‌های کم تراکم، با افزایش فاصلهٔ بین جمعیت مرجع و جمعیت‌های آزمون، نرخ درستی اسناد کردن کاهش یافت. روش‌های اسناد کردن نژادگانی با شناسایی بخش‌های مشترک تک‌جوری بین افراد جمعیت مرجع و افراد جمعیت‌های آزمون، نژادگان‌های گمشده را پیش‌بینی می‌کنند. در حقیقت ایدهٔ ضمنی فرض شده در روش‌های اسناد کردن نژادگانی این است که نژادگان هر فرد در جمعیت آزمون، ترکیبی از ساختار تک‌جوری دیگر افراد است (Marchini & Howie, 2010). بنابراین می‌توان نتیجه گرفت که روش‌های اسناد کردن نژادگانی به ساختار نبود تعادل پیوستگی وابسته است. کاهش نرخ درستی اسناد کردن با افزایش فاصلهٔ بین جمعیت مرجع از جمعیت‌های آزمون (شکل ۳) ممکن است در نتیجهٔ رخداد نوترکیبی و از بین رفتن ساختارهای تک‌جوری با افزایش فاصله بین جمعیت مرجع و جمعیت‌های آزمون باشد. در این تحقیق درستی اسناد کردن با استفاده از تراشه‌های با تراکم متفاوت بررسی شد. با استفاده از اسناد کردن نژادگانی امکان پیش‌بینی و تلفیق اطلاعات نژادگانی تراشه‌های مختلف با تراکم‌های متفاوت امکان‌پذیر است. استفاده از نژادگان‌های اسنادشده افزون بر کاهش چشمگیر در هزینه‌های تعیین نژادگان سبب افزایش درستی انتخاب ژنگانی می‌شود.

**همبستگی بین ارزش‌های اصلاحی ژنگانی برآورد شده پیش و پس از اسناد کردن**

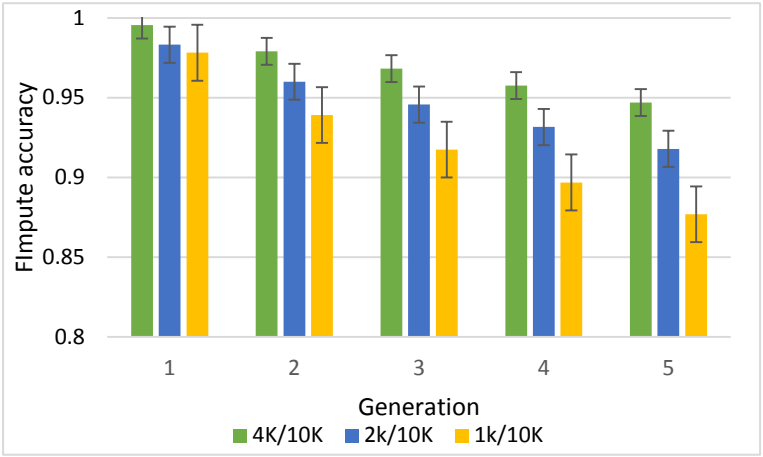
یکی از هدف‌های مهم در انتخاب ژنگانی برآورد درست ارزش‌های اصلاحی افراد است. در نتیجه بنا به‌ضرورت، تأثیر استفاده از تراشه‌های با تراکم مختلف و همچنین تأثیر استفاده از تراشه‌های اسنادشده روی تراشه با تراکم بالا در پیش‌بینی ارزش‌های اصلاحی افراد بررسی شد. ارزش‌های اصلاحی افراد با استفاده از تراشه‌های 1k، 2k و 4k برآورد شد. آنگاه همبستگی ارزش‌های اصلاحی برآورد شدهٔ ناشی از تراشه‌های با تراکم پایین و ارزش‌های اصلاحی برآورد شده با استفاده از تراشهٔ 10k محاسبه شد. همبستگی بین ارزش‌های اصلاحی برآورد شدهٔ به‌دست‌آمده از تراشهٔ 4k و تراشهٔ 10k بالاتر از همبستگی بین ارزش‌های اصلاحی برآورد شده با استفاده از تراشهٔ 10k و ارزش‌های اصلاحی برآورد شدهٔ به‌دست‌آمده از تراشه‌های 1k و 2k بود (شکل ۴). هنگامی‌که درستی ارزیابی‌های ژنگانی بالا باشد تا حدودی به همان نسبت، همبستگی بین ارزش‌های اصلاحی برآورد شده

www.SID.ir

Archive of SID

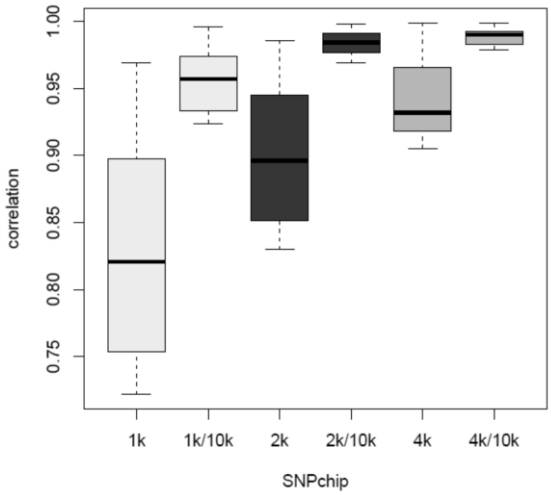


با استفاده از تراشه‌های با تراکم پایین و تراشهٔ پرتراکم بیشتر است. بنابراین با افزایش فاصلهٔ بین جمعیت مرجع و جمعیت‌های تأیید، درستی ارزیابی‌های ژنگانی کاهش‌یافته پس انتظار می‌رود همبستگی بین ارزش‌های اصلاحی برآورد شده با تراشهٔ پرتراکم و تراشه‌های با تراکم پایین کاهش می‌یابد. روند کاهشی بین همبستگی‌های برآورد شده برای تراشه 4k نسبت به تراشه‌های 1k و 2k کندتر بود. همچنین نتایج به‌دست‌آمده برای تراشه‌های اسنادشده روی تراشه با تراکم 10k همسان بود (شکل ۴). اما همبستگی‌های به‌دست‌آمده بیشتر بودند، به عبارتی پس از اسناد کردن می‌توان با درستی بالایی ارزش‌های اصلاحی افراد را برآورد کرد. افزایش میزان همبستگی برای تراشه‌های کم تراکم پس از اسناد کردن، بیشتر بود اما همچنان دامنهٔ تغییرپذیری درستی ارزش‌های اصلاحی برآورد شده در تراشه‌های کم تراکم بیشتر بود. این نتایج با تحقیق‌های گذشته همخوانی دارند (Szyda *et al*., 2013).




شکل ۳. درستی اسناد کردن با استفاده از FImpute در سه سطح 1K، 2K و 4K در پنج نسل جمعیت‌های تأیید یا آزمون

Figure 3. Imputation accuracy by using FImpute in three levels (1K, 2K and 4K) in five generations of validation or test populations




شکل ۴. همبستگی بین ارزش‌های اصلاحی برآوردشدهٔ ژنگانی در سه سطح (1k، 2k و 4k) و همبستگی بین ارزش‌های اصلاحی برآورد شده و تراشه‌های اسنادشده در سه سطح (1k/10k، 2k/10k و 4k/10k) و ارزش‌های اصلاحی برآوردشده با استفاده از تراشه با تراکم بالا 10k

Figure 4. Correlation between the estimated genomic breeding values in three levels (1k, 2k and 4k) and imputed panels in three levels (1k/10k, 2k/10k) and 4k/10k and estimated breeding values by using panel with 10k high density

www.SID.ir

Archive of SID



**نتیجه‌گیری کلی**

مقایسهٔ سه تراشه با تراکم متفاوت نشان داد، درستی ارزیابی‌های ژنگانی بر پایهٔ تراشهٔ 4k تفاوت بسیار کمی با نتایج تراشهٔ 10k دارد و همچنین نسبت به دو تراشهٔ دیگر 1k و 2k نتایج بهتری داشت. در این تحقیق اثر سه تراشهٔ 1k، 2k و 4k پیش و پس از اسناد شدن بر توان پیش‌بینی ارزش‌های اصلاحی افراد بررسی شد. به‌طورکلی با افزایش تراکم نشانگر توان پیش‌بینی ارزش‌های اصلاحی افراد حتی بدون اطلاع از ارزش‌های اصلاحی والدین افزایش می‌یابد. این نتیجه به‌نوبهٔ خود سهم زیادی در کاهش هزینه‌ها و افزایش منابع اطلاع اطلاعاتی افراد دارد. بنابراین با استفاده از تراشه‌های کم تراکم و آنگاه اسناد کردن آن‌ها به تراشه‌های با تراکم بالاتر می‌توان توان پیش‌بینی ارزش‌های اصلاحی افراد را افزایش داد.

## REFERENCES


1. Berry, D. P., McClure, M. C., Waters, S., Weld, R., Flynn, P., Creevey, C. & Mullen, M. P. (2013). Development of a custom genotyping panel for dairy and beef cattle breeding and research. *Advances in Animal Biosciences*, 4, 249.
2. Bolormaa, S., Gore, K., Werf, J. H. J., Hayes, B. J. & Daetwyler, H. D. (2015). Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. *Animal Genetics*, 46(5), 544-556.
3. Goddard, M. E. & Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, 124(6), 323-330.
4. Habier, D., Fernando, R. L. & Dekkers, J. C. (2009). Genomic selection using low-density marker panels. *Genetics*, 182(1), 343-353.
5. Harris, B. L., Creagh, F. E., Winkelman, A. M. & Johnson, D. L. (2011). Experiences with the Illumina high density bovine beadchip. *Interbull Bulletin*, (44).
6. Hayes, B. J. & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819-1829.
7. Judge, M. M., Kearney, J. F., McClure, M. C., Sleator, R. D. & Berry, D. P. (2016). Evaluation of developed low-density genotype panels for imputation to higher density in independent dairy and beef cattle populations. *Journal of Animal Science*, 94(3), 949-962.
8. Marchini, J. & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, 11(7), 499-511.
9. Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819-1829.
10. Ogawa, S., Matsuda, H., Taniguchi, Y., Watanabe, T., Takasuga, A., Sugimoto, Y. & Iwaisaki, H. (2016). Accuracy of imputation of single nucleotide polymorphism marker genotypes from low-density panels in Japanese Black cattle. *Journal of Animal Science*, 87(1), 3-12.
11. Sargolzaei, M., Chesnais, J. P. & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. BMC Genomics, 15:478.
12. Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123(4), 218-223.
13. Solberg, T. R., Sonesson, A. K. & Woolliams, J. A. (2008). Genomic selection using different marker types and densities. *Journal of Animal Science*, 86(10), 2447-2454.
14. Szyda, J., Zukowski, K., Kaminski, S. & Zarnecki, A. (2013). Testing different single nucleotide polymorphism selection strategies for prediction of genomic breeding values in dairy cattle based on low density panels. *Czech Journal of Animal Science*, 58(3), 136-145.
15. Technow FR [Package hypred]. (2011). Simulation of Genomic Data in Applied Genetics. University of Hohenheim.
16. VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91, 4414-23.


www.SID.ir

Archive of SID



17. VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F. & Schenkel, F. S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92, 16-24.
18. Wiggans, G. R., VanRaden, P. M. & Cooper, T. A. (2011). The genomic evaluation system in the United States: Past, present, future. *Journal of Dairy Science*, 94(6), 3202-3211.

www.SID.ir