Large Genomic Region Free of GWAS-Based Common Variants Contains Fertility-Related Genes

Rong Qiu, Chao Chen, Hong Jiang, Libing Shen, Min Wu, Chunyu Liu

School of Information Science and Engineering, Central South University, Changsha, China, Human Engineering Laboratory for Advanced Control and Intelligent Automation, Changsha, China, Department of Psychiatry, University of Illinois at Chicago, Chicago, United States of America, Institute of Human Genetics, University of Illinois at Chicago, Chicago, United States of America, Department of Neurology, Xiangya Hospital, Central South University, Changsha, China, School of Life Science, Fudan University, Shanghai, China, State Key Laboratory of Medical Genetics of China, Central South University, Changsha, China

Abstract

DNA variants, such as single nucleotide polymorphisms (SNPs) and copy number variants (CNVs), are unevenly distributed across the human genome. Currently, dbSNP contains more than 6 million human SNPs, and whole-genome genotyping arrays can assay more than 4 million of them simultaneously. In our study, we first questioned whether published genome-wide association studies (GWASs) assays cover all regions well in the genome. Using dbSNP build 135 data, we identified 50 genomic regions longer than 100 kb that do not contain any common SNPs, i.e., those with minor allele frequency (MAF) $\leq 1\%$. Secondly, because conserved regions are generally of functional importance, we tested genes in those large genomic regions without common SNPs. We found 97 genes and were enriched for reproduction function. In addition, we further filtered out regions with CNVs listed in the Database of Genomic Variants (DGV), segmental duplications from Human Genome Project and common variants identified by personal genome sequencing (UCSC). No region survived after those filtering. Our analysis suggests that, while there may not be many large genomic regions free of common variants, there are still some “holes” in the current human genomic map for common SNPs. Because GWAS only focused on common SNPs, interpretation of GWAS results should take this limitation into account. Particularly, two recent GWAS of fertility may be incomplete due to the map deficit. Additional SNP discovery efforts should pay close attention to these regions.

Introduction

The human genome contains millions of common SNPs, which are being deposited into public databases. These data have been used to design genome-wide association studies (GWASs) [1,2,3]. Common SNPs are better powered in association tests [4]. However, genomic regions not covered by common variants are neglected. Those neglected regions may contain variants with low frequencies, and should be paid more attention to because rare variants are even more likely to be functional than common ones [5].

In our study, we were interested in two questions: 1) whether the human genome is sufficiently covered by common SNPs and is sufficiently captured by common SNPs of standard GWAS platforms, and 2) whether any genes were included in those regions and their enriched biological functions.

To answer these two questions, we started with searching regions without common SNPs, called common SNP-free regions (CSFRs), regions free of both common SNPs and CNVs, called common variant-free regions (CVFRs). Next, we explored the functional enrichment of genes identified in CSFRs and CVFRs. With available personal genome sequencing data, whether these CSFRs and CVFRs contain common and rare variants were also examined.

Methods

Identification of CSFRs and CVFRs

Common SNPs (MAF \(\geq 1\%\)) in dbSNP build 135, Genome Assembly Gaps and Genome Database refGene data were downloaded from the UCSC Genome Browser (http://genome.ucsc.edu/) (Table 1). The CNV data were downloaded from the DGV (Table 1). Using the common SNP table, we calculated distances between adjacent common SNPs and subtracted regions containing the genome assembly gaps. If the remaining SNP intervals were longer than 100 kb, those intervals were defined as CSFRs. The CSFRs were further searched for CNVs. If after subtracting regions containing CNVs, the intervals were still longer than 100 kb, those intervals were defined as CVFRs. The reason we used 100 kb as bin to detect SNP free region is the SNP linkage disequilibrium distance: several groups reported blocks of up to 100 kb in length exhibiting very strong linkage disequilibrium [6,7].

To verify our result for its impacts on GWAS, we first determined whether the CSFRs are truly missed by Affymetrix
Pathway and functional analyses

Identification of genes in CSFRs and CVFRs

Gene annotation data from the Human Genome assembly hg19 UCSC refGene was used to map coding genes in the CSFRs and CVFRs (Table 2). Genes were included if their transcription regions overlapped with the CSFRs/CVFRs by at least one base pair. When a gene had multiple splicing forms, we chose the longest splicing form to define the gene region.

Pathway and functional analyses

The genes identified in the CSFRs/CVFRs were used to analyze their enrichment of biological functions through the Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/tools.jsp).

Isochore characterization

Isochore is a large region of DNA sequence which has a relatively uniform degree in its GC content [8]. We use 100 kb as the length of flank region and 2% GC difference as indicator to identify isochore, isochore border and unknown region among SNP free regions. All SNP free regions in this study are longer than 100 kb. CSFRs are identified as isochore if its GC content is 2% greater or lower than both right and left regions. CSFRs are identified as isochore border if the difference of GC content between two flank regions is greater than 2%, and GC-content difference between left flank and right flank region is greater than GC-content difference between CSFR and its flank regions. Unknown region means CSFR is neither isochore nor isochore border.

Table 1. Data Sources Used in This Study.

| Data                                      | URL                                                                 | Version               | Modified date      | Data description and summary statistics                                                                 |
|-------------------------------------------|---------------------------------------------------------------------|-----------------------|--------------------|--------------------------------------------------------------------------------------------------------|
| Common SNP Data in HapMap                 | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database              | Human Genome assembly 18-Dec-2011 hg19. |                       | snp135Common.txt.gz Total SNPs: 11,488,259 in chr1-chrY.                                           |
| Genome Assembly Gaps data                 | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database              | Human Genome assembly 27-Apr-2009 hg19. |                       | gap.txt.gz Total gaps, 357 in chr1-chrY.                                                            |
| Genomes Unzipped data                     | http://www.genomesunzipped.org/download/                             | Based on Human genome 10-Oct-2010 hg18, upgraded to hg19. |                       | Total of 1923 SNPs in the chrY sample, 546 common SNPs with maf>1%.With data for 9 personal genome sequences. |
| personal genome variation data            | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/             |                       |                     | Total of 9 personal genomes: pgNA12687.txt.gz, pgNA12891.txt, pgNA12892.txt.gz, pgNA19240.txt, pgNA9340.txt, pgCH2010_Human_Genome_genome.txt.gz, pgX11.txt.gz, pgY1.txt.gz. |
| DGV data                                 | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/             | Human Genome assembly 07-Mar-2011 hg19. |                       | dgv.txt.gz Total 101605 in chr1-chrY.                                                               |
| segmental duplication data                | http://echlerlab.gs.washington.edu/database.html                    | Human Genome assembly 27-Jun-2011 hg19. |                       | inter pairs is 22980; intra pairs is 8763                                                          |
| Genes                                     | http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz | Human Genome assembly 21-May-2012 hg19. |                       | Total number of genes is 42,742; after eliminating other chromosome, 30,332 genes in chr1-chrY remain. |

Results

CSFRs and CVFRs identification

We identified 50 CSFRs distributed across eight chromosomes: chr1, chr2, chr7, chr9, chr10, chrX, and chrY. The Y chromosome carried the majority of these regions—33 in total (Table 2). After excluding the CNV regions, we identified 20 CVFRs distributed across two chromosomes: chrX and chrY. The Y chromosome still carried the majority, with 18 regions (Table 3).

We checked our results in the Affymetrix SNP Array 6.0 by its annotation data. Among the CSFRs, we found 25 SNPs’ information in the annotation file, and only four of them had non-zero minor allele frequency: rs11681529, rs2571764, rs2874557, and rs35516764. The other 20 are monomorphic for the HapMap four populations (Caucasian, African, Chinese and Japanese). Therefore, we concluded that most of these 50 large genomic regions has not been covered properly by the Affymetrix 6.0 Array at least in those major populations investigated.

Genes in CSFRs and CVFRs and their functional enrichment

Ninety-seven genes overlapped with 28 of the 50 CSFRs (56%) (Table 2). DAVID was used to test whether the annotations of this set of genes were over presented with particular GO terms [9]. They were highly enriched with biological pathways involved with sexual reproduction, spermatogenesis, male gamete generation, gamete generation, multicellular organism reproduction, and reproductive processes in a multicellular organism (p<0.05 and FDR q<0.05, Table 4). The gene set included a number of gene previously reported to be related to reproduction, including DAZL [10,11], BPI2 [12], TSP12 [11], CD17 [13], CD124 [13] and RBMY1 [11]. A gr/gr deletion polymorphism on Y chromosome still carried the majority, with 18 regions (Table 3).

Twenty genes were overlapped with seven of the 20 CVFRs (35%) (Table 3). DAVID was also performed on these 20 genes. However, these genes were not enriched in any biological functions.
Table 2. List of 50 common SNP-free regions containing 97 genes.

| Chr | CSFR_start    | CSFR_end     | CSFR_size | Gene_name                      | Isochore_type          |
|-----|---------------|--------------|-----------|-------------------------------|------------------------|
| chr1| 145883118     | 145989503    | 106385    | GPR89C, PDZK1P1               | Isochore_border        |
| chr1| 111191098     | 111347035    | 155937    | LIM53-LIM53-LOC440895, LIM53 | Isochore               |
| chr2| 110524226     | 1179805      | 742528    | RGPD5, RGPD6, LIM53, LIM53    | Isochore_border        |
| chr2| 41497718      | 4135419      | 123701    | FAM75A5, FAM75A7, LOC653501, | Unknown                |
| chr2| 42743905      | 4284739      | 103489    | LOC286297                     | Isochore_border        |
| chr10| 46799214     | 4690775      | 108561    | FAM35B                         | Isochore               |
| chr7 | 74765724      | 7466460      | 100736    | GAT5L2                         | Isochore_border        |
| chr9 | 39379250      | 3955146      | 172206    | LOC653501, ZNF658B            | Unknown                |
| chr9 | 39829606      | 3996180      | 132198    | FAM75A2, FAM75A1, FAM75A1    | Unknown                |
| chr9 | 41497718      | 4135419      | 123701    | FAM75A5, FAM75A7, LOC653501, | Unknown                |
| chr9 | 42743905      | 4284739      | 103489    | LOC286297                     | Isochore_border        |
| chr10| 48185336     | 48300420     | 115084    | LOC642826, AGAP9, FAM25B,    | Isochore_border        |
| chr16| 33142890      | 3329378      | 150888    | TP53TG3, TP53TG3C, TP53TG3B  | Isochore_border        |
| chrX | 52098738      | 5239514      | 297176    | XAGE2, XAGE2B, XAGE1B, XAGE1A, | Unknown                |
| chrX | 52445914      | 5256230      | 122316    | XAGE1A, XAGE1C, XAGE1E, XAGE1D, | Isochore_border        |
| chrY | 4834281       | 4935713      | 101432    | PCDH11Y                       | Isochore_border        |
| chrY | 5012892       | 5025540      | 192648    | PCDH11Y                       | Unknown                |
| chrY | 5274434       | 5421065      | 146631    | PCDH11Y                       | Isochore_border        |
| chrY | 6074690       | 642524       | 347834    | TTTY23, TTTY23B, TTTY1B,    | Isochore               |
| chrY | 9381846       | 9492957      | 111111    | RBMY3AP                       | Isochore               |
| chrY | 9524503       | 9788115      | 243612    | TTTY8, TTTY8B, TTTY7, TTTY7, | Isochore_border        |
| chrY | 14691127      | 14804076     | 112949    | TTTY15                         | Isochore_border        |
| chrY | 19563894      | 20143885     | 579991    | FAM41AY1, FAM41AY2, LINC00230B, | Unknown                |
| chrY | 20193885      | 20834702     | 640817    | XRY, XRY2, LINC00230A, LINC00230A, | Unknown                |
| chrY | 20837553      | 21080706     | 243153    | TTTY9B, TTTY9A, HSFY2, HSFY1, | Isochore_border        |
| chrY | 22564778      | 22665261     | 100483    | TTTY10                         | Unknown                |
| chrY | 23473201      | 23580342     | 107141    | RBMY2EP                       | Isochore_border        |
| chrY | 23634362      | 23838234     | 203872    | RBMY1B, RBMY1A1, RBMY1E, RBMY1D, | Isochore_border        |
| chrY | 23993156      | 24359930     | 366774    | RBMY1A1, RBMY1D, RBMY1B, RBMY1E, | Isochore_border        |
| chrY | 24500502      | 24620459     | 119857    | RBMY1F, RBMY1J, TTTY6, TTTY6  | Unknown                |
| chrY | 24620459      | 28160890     | 354043    | PRY, PRY2, TTTY1B, TTTY1C,TTTY1A, | Isochore_border        |
| chr9 | 42027732      | 42145811     | 118079    | Isochore_border               |                       |
| chr9 | 44466205      | 44651655     | 185450    | Isochore_border               |                       |
| chr9 | 45128500      | 45250203     | 121703    | Isochore_border               |                       |
| chr9 | 65632583      | 65745692     | 113109    | Isochore_border               |                       |
| chrY | 3179117       | 3359419      | 180302    | Isochore_border               |                       |
| chrY | 3833777       | 3966707      | 132930    | Isochore_border               |                       |
SNP-free regions from personal genome sequencing and segmental duplications

We further explored those SNP-free regions in personal genome variant data. Rare variants were detected in most of the CSFRs or CVFRs. Only one region on X chromosome (chrX: 52,267,361-52,395,914) left. We also examined this region in updated dbSNP database (dbSNP137, http://www.ncbi.nlm.nih.gov/). Two more common SNPs were detected (rs201652812 and rs199865557). After subtract them, the left region was 105 kb (chrX: 52,290,698-52,395,914), which was the finally region not containing any

Table 3. List of 20 common variant-free regions containing 20 genes.

| chr  | CVFR_start | CVFR_end | CVFR_size | gene_name     |
|------|------------|----------|-----------|---------------|
| chrX | 3966708    | 4346934  | 380226    | XAGE2, XAGE2B |
| chrX | 4466077    | 4593373  | 122796    | Unknown       |
| chrY | 4593411    | 4807708  | 214297    | Unknown       |
| chrY | 6482140    | 6677618  | 195478    | Unknown       |
| chrY | 7401836    | 7548914  | 147078    | Unknown       |
| chrY | 8214827    | 8334874  | 120047    | Isochore_border |
| chrY | 15039955   | 15234829 | 194874    | Unknown       |
| chrY | 18248698   | 18381734 | 133036    | Unknown       |
| chrY | 18390543   | 18560004 | 169461    | Isochore_border |
| chrY | 19357294   | 19500106 | 124812    | Unknown       |
| chrY | 22214221   | 22369679 | 155458    | Isochore_border |
| chrY | 22419679   | 22564743 | 145064    | Isochore_border |
| chrY | 23241568   | 23361665 | 120097    | Isochore_border |
| chrY | 28160891   | 28509481 | 348590    | Isochore_border |
| chrY | 15039955   | 15234829 | 194874    | Unknown       |
| chrY | 22214221   | 22369679 | 155458    | Isochore_border |
| chrY | 22419679   | 22564743 | 145064    | Isochore_border |
| chrY | 23241568   | 23361665 | 120097    | Isochore_border |
| chrY | 28160891   | 28509481 | 348590    | Isochore_border |
| chrY | 18248698   | 18381734 | 133036    | Unknown       |
| chrY | 18390543   | 18560004 | 169461    | Isochore_border |
| chrY | 19357294   | 19500106 | 124812    | Unknown       |
| chrY | 22214221   | 22369679 | 155458    | Isochore_border |
| chrY | 22419679   | 22564743 | 145064    | Isochore_border |
| chrY | 23241568   | 23361665 | 120097    | Isochore_border |
| chrY | 28160891   | 28509481 | 348590    | Isochore_border |
| chrY | 15039955   | 15234829 | 194874    | Unknown       |
| chrY | 22214221   | 22369679 | 155458    | Isochore_border |
| chrY | 22419679   | 22564743 | 145064    | Isochore_border |
| chrY | 23241568   | 23361665 | 120097    | Isochore_border |
| chrY | 28160891   | 28509481 | 348590    | Isochore_border |

doi:10.1371/journal.pone.0061917.t003
Table 4. Top 6 GO terms from the functional annotation analysis of 97 CSFR genes by DAVID.

| Category | Term | Count | % | P-Value | FDR |
|----------|------|-------|---|---------|-----|
| GOTERM_BP_FAT | sexual reproduction | 9 | 14.8 | 0.00000003 | 0.000033 |
| GOTERM_BP_FAT | Spermatogenesis | 8 | 13.1 | 0.00000047 | 0.000052 |
| GOTERM_BP_FAT | male gamete generation | 8 | 13.1 | 0.00000047 | 0.000052 |
| GOTERM_BP_FAT | gamete generation | 8 | 13.1 | 0.00000026 | 0.000028 |
| GOTERM_BP_FAT | multicellular organism reproduction | 8 | 13.1 | 0.0000011 | 0.0012 |
| GOTERM_BP_FAT | reproductive process in a multicellular organism | 8 | 13.1 | 0.0000011 | 0.0012 |

1gene included RBMY1A1, RBMY1B, RBMY1J, RBMY1F, XKRY, XKRY2, BPY2C, BPY2B, BPY2, CDY1, CDY1B, CDY2B, CDY2A, CDY1B, CDY2B, CDY2A, DA2Z, DA2Z, DA2Z, DA2Z, DA2Z, and TSPY2.
2gene included RBMY1A1, RBMY1B, RBMY1J, RBMY1F, BPY2C, BPY2B, BPY2, CDY1, CDY1B, CDY2B, CDY2A, DA2Z, DA2Z, DA2Z, DA2Z, DA2Z, and TSPY2.

Discussion

We performed a thorough search for large genomic regions that are free of common variants in dbSNP and we found 50 CSFRs and 20 CVFRs. Most of these variations free regions located on Y chromosome. Genes in the CSFRs were highly enriched for activities related to reproduction. Further investigation in the sequencing of personal genomes found most of the CSFRs (49 out of 50) did contain rare SNPs, suggesting those regions have not been covered well in the existing common variants sequencing projects, like the 100 Genomes Project.

GWAS is one the most infusive common variants sequencing projects, but important finding might be missed because of its poor coverage of rare variants. Recently, two fertility GWAS studies were conducted but failed to find SNPs on sex chromosomes [16,17]. Both studies used Affymetrix GWAS platforms that we evaluated in this study. However, both sex chromosomes have long been implicated in infertility, specifically in spermatogenic damage in mouse models and in human candidate gene/region studies [18]. Our study found that those genomic regions free of common variants regions carrying many genes important to reproduction. With those important candidate genes missing, we must be cautious of analyzing fertility-related GWASs, which may produce false negatives.

The most reliable CVFR call contains the XAGE2 and its isoforms, which belong to XAGE subfamily. XAGE2 is strongly expressed in normal testes, and in some tumor [19]. Because genotyping platforms cannot fully cover structural variations such as segmental duplication, we further applied structural variations filtering analysis, and observed XAGE region was overlapped with segmental duplication. Based on these observations, we concluded that the observation of variant free regions is more a coverage problem with the current versions of dbSNP and existing GWAS assay platforms than a lack of assavable variation. When more genomes are sequenced, we may end up with proper coverage of complete human genome by common SNPs.

We mapped our SNPs on dbSNP build 135 and regions on GRC37.p10 (hg19) assembly reference, which is the most accurate alignment version and with all current genome knowledge available. Comparing to old versions, hg19 changed many genomic coordinates and included alternate haplotype assemblies for chr6 (7 haplotypes), chr4 (1 haplotype), and chr17 (1 haplotype). Different versions can be converted by liftOver software (http://genome.ucsc.edu/cgi-bin/hgLiftOver). More details of differences in each version are provided in NCBI (http://www.ncbi.nlm.nih.gov/genome/guide/human/release_notes.html).

Further study can focus on the sequence properties of those regions, and their conservative across species. Isochores are spatially heterogeneous in mammalian genome and varies in replication timing, gene richness, recombination rate, etc [20,21,22]. Natural selection is the most plausible explanation for formation and maintenance of isochores [20]. We observed nearly half of CSFRs are isochores and isochore border regions, which is a hint that these CSFRs may be under different selection pressure from its neighboring regions. To further test selection pressure, we mapped those regions to chimpanzee and mouse by Synteny analysis from Ensembl (http://useast.ensembl.org/Homo_sapiens/Location/Synteny?r = 6:133017695-133161157), and found only 6 genes (RGPD5, RGPD6, GATS2L, FAAM2G, HSFY1, HSFY2) can map to unique regions in the other two species. Next we applied dN/dS ratio test, the ratio of substitution rates at non-synonymous and synonymous sites, and found that human genes under more purify selection than chimpanzee genes (paired T test, p = 0.01, Table S2). Those results suggest that natural selection seems to be the major evolutionary force behind these variant-free regions.

In summary, by searching large genomic regions free of common variants for the first time, we identified tens of common variations free regions, and most of them were located on the X and Y chromosomes. The genes located in CSFRs are enriched for fertility. Incorporating personal genome data, only one region was still free of variants and harbored gene XAGE2, indicating most of
the detections due to low coverage of rare variations. Future deep sequencing from more individuals and redesigning GWAS arrays should improve our understanding of the variability of these regions and their functional importance.

Supporting Information

Table S1  Isochore characterization of 50 CSFRs. (DOC)

Table S2  Evolution pressure of conserved genes by dN/dS ratio test. (DOC)

Author Contributions

Conceived and designed the experiments: CL. Performed the experiments: RQ CC. Analyzed the data: RQ LS. Contributed reagents/materials/analysis tools: CC MW HJ. Wrote the paper: RQ CC HJ.

References

1. Jiang RH, Duan JC, Windemuth A, Stephens JC, Judson R, et al. (2003) Genome-wide evaluation of the public SNP databases. Pharmacogenomics 4: 779–789.
2. Wang WYS, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: Theoretical and practical concerns. Nature Reviews Genetics 6: 109–118.
3. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews Genetics 9: 356–369.
4. Nannya Y, Taura K, Kurokawa M, Chiba S, Ogawa S (2007) Evaluation of genome-wide power of genetic association studies based on empirical data from the HapMap project. Human Molecular Genetics 16: 2494–2505.
5. Zhu QQ, Ge DL, Maia JM, Zhu MF, Petrovski S, et al. (2011) A Genome-wide Comparison of the Functional Properties of Rare and Common Genetic Variants in Humans. American Journal of Human Genetics 88: 458–468.
6. Aissani B, Perusse L, Lapointe G, Chagnon YC, Bouchard L, et al. (2006) A quantitative trait locus for body fat on chromosome 1q43 in French Canadians: linkage and association studies. Obesity (Silver Spring) 14: 1605–1615.
7. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE (2001) Segmental duplications: Organization and impact within the current Human Genome Project assembly. Genome Research 11: 1005–1017.
8. Costantini M, Clay O, Auleta F, Bernardi G (2006) An isochore map of human chromosomes. Genome Res 16: 536–541.
9. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44–57.
10. Fernandes S, Huellen K, Goncalves J, Dukal H, Zeiler J, et al. (2002) High frequency of DAZ1/DAZ2 gene deletions in patients with severe oligozoospermia. Mol Hum Reprod 8: 286–298.
11. Lardone MC, Parodi DA, Valdeveneto R, Ehrenpeger M, Piantante A, et al. (2007) Quantification of DDX5Y, RBMI1, DAZ and TSPY mRNAs in testes of patients with severe impairment of spermatogenesis. Mol Hum Reprod 13: 705–712.
12. Choi J, Koh E, Suzuki H, Maeda Y, Yoshida A, et al. (2007) Alu sequence variants of the BPT2 gene in proven fertile and infertile men with Sertoli cell-only phenotype. Int J Urol 14: 431–435.
13. Kleiman SE, Lehavi O, Rauker A, Botchan A, Paz G, et al. (2011) CDY1 and BOU/LE transcripts assessed in the same biopsy as predictive markers for successful testicular sperm retrieval. Fertil Steril 95: 2297–2302, 2292 e2291.
14. Repping S, Skalenky H, Brown L, van Daalen SK, Kover CM, et al. (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haplotype selection. Nat Genet 35: 247–251.
15. Krausz C, Giachini C (2007) Genetic risk factors in male infertility. Arch Androl 53: 125–133.
16. Kosova G, Scott NM, Niederberger C, Prins GS, Ober C (2012) Genome-wide association study identifies candidate genes for male fertility traits in humans. Am J Hum Genet 90: 950–961.
17. Hu ZB, Xia YK, Guo XJ, Dai JC, Li HG, et al. (2012) A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. Nature Genetics 44: 103–106.
18. Burgoyne PS, Mahadevaiah SK, Satchell MJ, Palmer SJ (1992) Fertility in Mice Requires X-Y Pairing and a Y-Chromosomal Spermiation Gene-Mapping to the Long Arm. Cell 71: 391–398.
19. Chen YT, Ross DS, Chin R, Zhou XK, Chen YY, et al. (2011) Multiple Cancer/Testis Antigens Are Preferentially Expressed in Hormone-Receptor Negative and High-Grade Breast Cancers. Plos One 6.
20. Costantini M, Cammarano R, Bernardi G (2009) The evolution of isochore patterns in vertebrate genomes. BMC Genomics 10: 146.
21. Oliver JL, Cappella P, Hackenberg M, Bernaola-Galvan P (2004) IsoFinder: computational prediction of isochores in genome sequences. Nucleic Acids Res 32: W287–292.
22. McVeany GA, Myers SR, Hunt S, Deloukas P, Bentley DR, et al. (2004) The fine-scale structure of recombination rate variation in the human genome. Science 304: 581–584.