# 

**UNIVERSIDADE DE BRASÍLIA**

**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA E BIODIVERSIDADE**

**REDE EM ASSOCIAÇÃO AMPLA**

**EDITAL Nº 01/2015**

**DETECÇÃO E ANÁLISE DE MUTAÇÕES DE NOVO EM PACIENTES COM EXPOSIÇÃO PARENTAL À RADIAÇÃO IONIZANTE DE CÉSIO-137 A PARTIR DE DADOS DE GENOTIPAGEM DE POLIMORFISMOS DE BASE ÚNICA DE ALTA DENSIDADE**

### Nome do Candidato: Hugo Pereira Leite Filho

E-mail: filho.hugo@gmail.com

Telefone: 62-999047721

### 27/11/2017

### Goiânia-GO Universidade Federal de Goiás - UFGo

1. **INTRODUÇÃO**

A capacidade para predição genética nas exposições humanas sob efeitos de radiação ionizantes vem sendo objeto de pesquisa ao longo de 50 anos. A radiação ionizante (IR) é um agente mutagênico amplamente estudada, a sua exposição resulta em diferentes tipos de lesões no DNA, quem vão desde as mudanças em nucleotídeos, até o rompimento de fita da cadeia dupla de DNA. Desta forma, prevalence a relevância biológica de cinco tipos de lesões no DNA induzidos por RI para estudos cinéticos, sendo que foram realizado sem células eucarióticas por vários laboratórios. Estas lesões foram identificadas em ligações cruzadas de DNA-proteína, dano base, quebras de cadeia simples, quebras de cadeia dupla e lesões volumosas (Adewoye et al., 2015; Frankenberg-Schwager, 1990).

Os efeitos mutagênicos de IR sobre a linha germinal são particularmente preocupantes como levam ao acúmulo de mutações adicionais em filhos de pais irradiados. Apesar dos numerosos esforços, pouco se sabe sobre os efeitos genéticos da exposição à radiação em seres humanos e a única evidência definitiva para a indução de mutação germinativa *in vivo* em mamíferos são retratados de estudos com ratos (UNSCEAR, 2001; Nakamura et al., 2013).

Desde o acidente em 1987 ocorrido em Goiânia, com o material radioativo Césio-137, até os dias de hoje, têm sido realizados vários estudos sobre a saúde genética dos radioacidentados goianos. Um dos primeiros testes de biomonitoramento das populações expostas à RI do Césio-137 foi o teste de micronúcleo, que relatou um aumento na frequência de micronúcleos das pessoas envolvidas direta ou indiretamente no acidente (da Cruz et al., 1994).

Atualmente foi estabelecido que a análise cromossômica pelo microarranjo (CMA – do inglês, *Chromosomal Microarray Analysis*) consiste em uma técnica de citogenética molecular, desenvolvida para a detecção de deleções ou duplicações de um amplo espectro de regiões clinicamente significativas do genoma humano (Lu *et al.,* 2008; Kong *et al.,* 2013). O CMA é capaz de detectar erros genômicos em regiões críticas dos cromossomos, geralmente não detectados através da investigação citogenética tradicional. Utilizam-se, na rotina do CMA, chips de DNA e hibridização genômica comparativa (CGH), sendo investigadas para aumento ou diminuição da dosagem várias regiões cromossômicas relevantes, permitindo detectar variações no número de cópias (CNVs) (Coulter*et al.,* 2011).

Os Microarranjos é uma poderosa ferramenta de diagnóstico que pode gerar uma gama de informação considerado sobre gene retirado de uma célula em uma só vez. Vários algoritmos de classificação podem ser aplicados na técnica de microarranjo para o desenvolvimento de métodos que possam predizer a ocorrência de uma doença. No entanto, a precisão de tais métodos difere de acordo com o algoritmo de classificação aplicada. Identificar o melhor algoritmo de classificação se torna um grande desafio (Mudunuri*et al*., 2011).

Xu et al (2011) afirma que as matrizes de genotipagem, provenientes de estudos de associação do genoma inteiro, do inglês *Genome Wide Association Studies* (GWAS) foram desenvolvidas para caracterizar o polimorfismo de nucleotídeo único (SNP – do inglês *Single-nucleotidepolymorphisms*) e DNA com variação de número de cópias (CNVs – do inglês *copy number variations*). Os algoritmos estatísticos não parametrizados e baseados em modelos foram desenvolvidos para detectar CNVs a partir de SNP, utilizando as intensidades de marcadores. No entanto, esses algoritmos necessitam de especificidade para detectar pequenas CNVs devido à elevada taxa de falsos positivos quando se chama CNVs com base nos valores de intensidade. Portanto, os testes de associações resultante da falta de CNVs afetam o diagnóstico de doenças comuns.

GWAS são estudos de “caso-controle” que examina SNPs para determinar fatores genéticos com doenças complexas. Nas últimas décadas, GWAS vem sendo utilizados em muitas doenças comuns de diferentes natureza, incluindo câncer, doenças cardiovasculares, doenças infecciosas, doenças psiquiátricas como por exemplo o autismo (Li, *et al.* 2014; Smoller, *et al*. 2013).

Dados originados de GWAS podem ser aplicados para identificação de CNVs. Assim, discutiremos alguns métodos para utilização de dados de SNPs para detectar eventos no número de cópias. Serão revisados uma série de algoritmos projetados para detectar as variações de número de cópias, por meio da utilização de dados com apresentações de sinais de intensidades, implementando aplicações em modelos estatísticos. Assim, será possível a realização de comparações de métodos para avaliar a precisão e detecção das alterações de CNVs.

1. **JUSTIFICATIVA**

No ano de 1987, em Goiânia-Goiás (Brasil), uma série de eventos inesperados resultou em um grave acidente radiológico, provocando um rastro de contaminação humana, animal, vegetal e ambiental por um radionuclídeo.

Os efeitos mutagênicos da radiação ionizante (RI) sobre a linhagem germinativa são de preocupação especial, pois podem perdurar por várias gerações, levando ao acúmulo de mutações adicionais em filhos de pais irradiados.

Várias técnicas moleculares têm permitido avanços consideráveis no campo da citogenética clínica e superaram as limitações do cariótipo convencional, visando detectar rearranjos não identificados na microscopia óptica que permitem avaliar pequenas perdas e ganhos de material genético (Yiping*et al*. 2010; Kong *et al.*, 2013; Shi*et al.*, 2013).

Para Jiang et al (2010), estudar as variações genéticas relacionadas às doenças, usando os estudos sobre GWAS vem sendo bastante aplicado em ensaios com base em centenas de milhares de SNPs, com objetivo de descobrir de forma individual ou em combinação de SNPs associados aos achados em doenças. O método comum de análise é empregar estatísticas parametrizadas com ajustes de testes realizados para limitar os achados falsos-positivos.

A pesquisa genômica pode levar a descobertas de variações no número de cópias (CNVs) que pode ser um fator de relevância aos irradiados. Estabelecer um método de identificação de CNVs por microarranjos de alta densidade, podendo contribuir com o aumento do conhecimento sobre os efeitos biológicos de exposição humana à radiação ionizante, possibilitando estimar a frequência de mutações germinativas na prole de indivíduos expostos acidentalmente à radiação ionizante de Césio-137.

A pesquisa trata sobre o grupo constituído por 12 famílias, dos quais pelo menos um dos progenitores foi diretamente exposto à radiação ionizante de Césio-137, incluindo um total de 40 indivíduos (12 casais e 16 filhos nascidos após o acidente). A dose absorvida para os indivíduos expostos foi estimada em ≤0,2 Gray. Um grupo de indivíduos não-expostos à radiação ionizante foi usado como controle. Esse grupo foi composto por 8 famílias goianas sem histórico de exposição à RI.

O estudo propõe examinar modelos de algoritmos aplicados na detecção e predição de CNVs em matrizes de genotipagem de SNPs aos indivíduos, e suas proles, que foram expostos acidentalmente à radiação ionizante de Césio-137.

1. **OBJETIVOS**
   1. **Objetivo Geral**

  Realizar um estudo com base nos conhecimentos da Bioinformática que usando informações de SNP’s dos trios buscará identificar a origem cromossômica de desvios mendelianos, contribuindo com o aumento do conhecimento sobre os efeitos biológicos de exposição humana à radiação ionizante, possibilitando estimar a frequência de mutações germinativas na prole de indivíduos expostos acidentalmente à radiação ionizante de Césio-137.

* 1. **Objetivos Específicos**
* Analisar as CNV’s mutáveis e não mutáveis de trios no intuito de identificar as causas de desvios mendelianos nas proles;
* Realizar um estudo sobre métodos de busca, por meio de algoritmos, aos achados de SNP’s;
* Comparar e analisar as ferramentas computacionais de investigação;
* Mapear um fluxo para auxiliar na compreensão da metodologia;
* Identificar o melhor modelo estatístico por meio de algoritmos de aprendizagem;
* Manter a continuidade aos estudos de monitoramento genético da população acidentalmente exposta ao Césio-137 em Goiânia-Goiás, contribuindo com o conhecimento sobre os efeitos biológicos de exposição à radiação ionizante.

1. **METODOLOGIA**
   1. **Grupo amostral**

O grupo exposto foi constituído por 12 famílias, dos quais pelo menos um dos progenitors foi diretamente exposto à radiação ionizante de Césio-137, incluindo um total de 40 indivíduos (12/12 pais/mães e 16 filhos nascidos após o acidente). A dose absorvida para os indivíduos expostos foi estimada em ≤0,2Gy.Um grupo de indivíduos não-expostos à radiação ionizante foi usado como controle. Esse grupo foi composto por 8 famílias goianas sem histórico de exposição à RI. As amostras biológicas foram constituídas de sangue periférico, sendo coletados 10 mL porpunção venosa. Para cada amostra biológica, foram separadas alíquotas de plasma, anel leucocitário e hemácias, sendo armazenadas a -20ºC. O anel leucocitário foi usado subsequentemente para extração e purificação do DNA genômico. O material restante foi armazenado para estudos futuros nos termos da Resolução CNS Nº 441/11.

No momento da coleta, os participantes responderam voluntariamente a um questionário e assinaram a um Termo de Consentimento Livre e Esclarecido. Todas as coletas foram procedidas nos domicílios dos participantes dos grupos de radioexpostos. Os controles doaram suas amostras voluntariamente no laboratório de genética do Núcleo de Pesquisas Replicon da PUC Goiás.

O presente estudo foi aprovado pelo comitê de ética e pesquisa da Pontifícia Universidade Católica de Goiás (PUC-Goiás) e o número do CAAE foi 49338615.2.0000.0037.

* 1. **Extração e Quantificação das Amostras de DNA:**

Para a extração de DNA foi utilizado o Kit de extração de DNA IllustraBloodGenomicPrep® Mini Kit (GE Healthcare, EUA). A quantificação da concentração de DNA (ng/µL) existente em cada amostra foi realizada em um espectrofotômetro NanoVue® Plus (GE Healthcare Life Sciences, Reino Unido). Ambos os procedimentos foram executados de acordo com os protocolos sugeridos pelos fabricantes.

* 1. **Análise Cromossômicas por Microarranjos**

A análise cromossômica por microarranjos (CMA) foi conduzida em um GeneChipCytoScan HD® (Affymetrix – Estados Unidos da América, EUA), sendo uma matriz de genotipagem abrangente para o genoma humano. A matriz de genotipagem apresenta ampla cobertura do genoma e maior desempenho para a análise de alterações cromossômicas humanas, sendo capaz de detectar variações genéticas estruturais. O CytoScan HD® possui mais de 99% de sensibilidade na detecção de CNVs, determinação de perda de heterozigose (LOH) e baixos níveis de mosaicismo.

A matriz do chip possui cerca de 2,6 milhões de cópias de marcadores, incluindo aproximadamente 750 mil SNPs e cerca de 2 milhões de sondas não polimórficas. As sondas que integram o chip apresentavam tamanho de 25 pb.

* 1. **Tratamento e Análise dos Dados Obtidos em Matrizes na Genotipagem dos SNPs**

A grande quantidade de informações disponibilizadas pelo uso do GeneChip® HD faz com que sejam necessárias análises complexas para o tratamento dos dados.

A proposta visa aplicar os procedimentos de exame de Trio, ligando CNV para cada 3 indivíduo (Pai, Mãe e Filho).

* 1. **Técnicas Aplicadas na Utilização do GeneChip® HD**

Foi executada no Software ChAS® que possibilita investigar alterações estruturais ao longo do genoma, tais como variações no número de cópias, como duplicações e deleções de segmentos de DNA de regiões gênicas presentes nas amostras estudadas de modo comparativo, facilitando as interpretações dos resultados.

Foram fixados como filtros: 15 e 8 marcadores de SNPs para se detectar microduplicações e microdeleções, respectivamente, distribuídos com uma média ≤2.000 kb, limitando-se a fragmentos ≥1kb.

Para CytoScan® HD os marcadores são atribuídos individualmente a uma chamada do número de cópias (NC) através do “Modelo Oculto de Markov” (do inglês, Hidden Markov Models, HMM). As entradas específicas da amostra para o HMM em relação ao sinal de referência são Log2 ratio, que são convertidos em número de cópias (NC) para cada marcador.

O HMM usa o algoritmo Viterbi para calcular os conjuntos de marcadores entre cada número de cópias. Quando os marcadores são atribuídos entre os números pelo HMM, trechos contínuos de marcadores adjacentes ordenados por posição de cromossomo são agregados em segmentos. Esses segmentos são resultantes do número de cópia, número de marcadores no segmento e da posição do marcador que inicia e termina o segmento.

Sendo assim, os parâmetros predefinidos aplica um segmento de entrada de comprimento mínimo com 5 marcadores. (Affymetrix® Chromosome Analysis Suite 2.0 TM Software User Manual).

* 1. **Coleta e Processamento das Amostras**

Após assinatura do TCLE e a obtenção dos escores ADI-R, serão obtidas as amostras biológicas dos pacientes, correspondendo a 10 mL de sangue periférico. Posteriormente, parte de cada amostra será processada para a separação do anel leucocitário.

* 1. **SNP e Posições de Marcadores**

Ao se referir as posições individuais em um cromossomo, como posições de SNPs, pode se atribuir uma única coordenada. Existem diferentes convenções sobre considerar o primeiro par de bases de DNA no cromossomo como posição 0 (zero) ou posição 1 (um).

Para as posições de marcadores SNP’s, em regra geral aplicam a coordenada de posição de índice baseada em 1: arquivos CYCHP, arquivos CNCHP, páginas de detalhes da ferramenta NetAffx para marcadores SNP, páginas NCBI para posições SNP de entradas dbSNP e a Tabela de Gráficos no ChAS© Versão 1.0.

Considerando uma escolha aleatória de marcador SNP “S-2MOG4” gerada pela ferramenta ChAS. Este marcador possui seu correspondente atribuída de “rs7641618” na base de dados do dbSNP. No entanto, no site da NCBI está definida na posição do cromossomo 3 (ch3):186255049 em NCBI37 (hg19) ou chr3:187737743 em NCBI36 (hg18). No site da ferramenta NetAffx, corresponde o NCBI36 (hg18), na coordenada chr3:187737743.

Os valores das coordenadas são descritas nos arquivos do tipo CYCHP gerados pela ferramenta ChAS©.

* 1. **Métodos de detecção de SNP’s**

As matrizes de SNP foram inicialmente construídas para genotipar SNPs em milhares usando a técnica de GWAS simultaneamente. Os microarrays SNP mais conhecidos estão disponíveis a partir de fornecedores comerciais, como Illumina e Affymetrix (LaFramboise, 2009; Rincon, et al., 2011; Xu, et al. 2013).

De acordo com Caetano (2009), a identificação de SNPs distribuídos aleatoriamente pelo genoma está distribuída por meio do alinhamento de uma sequência de um fragmento aleatório do genoma com uma sequência consenso. Entretanto, existe outro método de identificação, sendo por meio de sequenciamento direto de fragmentos específicos do genoma amplificados por PCR, e subsequente alinhamento e comparação das sequências, desta forma aplica a mineração de SNPs em uma região de interesse.

* 1. **Workflow CytoScan®**

O fluxo de processo da ferramenta ChAS© para identificação de SNP’s e CNV’s começa com os dados de sinal bruto no arquivo CEL, sendo que essa ferramenta implementa uma série de etapas que realizam o resumo do conjunto de sondas, a normalização, a remoção da variação causada por propriedades conhecidas e a variação residual, e completando com chamadas de genótipos, segmentos de números de cópia e LOH segmentos.

**4.10 Algoritmos para detecção de SNP’s**

Estudos identificam aplicação de técnicas na detecção de SNP’s usando Algoritmo Genético (AG) para a otimização da busca e identificação de associações relevantes. O AG é um paradigma evolucionário (Goldberg, 1989), sendo que este algoritmo realiza buscas estocásticas baseado no processo da evolução Darwiniana a fim de encontrar soluções para problemas computacionalmente complexos. Os AGs são apropriados para estudos GWAS já que através de um processo aleatório, tal como seleção natural, mutação e cruzamento, atua na investigação de subconjunto de todas as possíveis interações.

Para Motsinger-Reif et al.,(2008) AG pode ser combinado com o algoritmo de colônia de formigas (Greene et al., 2008).

Estudos mapearam métodos que aplicam evolução gramatical de redes neurais - ‘*grammatical evolution*’ - utilizado para construir redes neurais (NN) e selecionar os SNPs associados à doenças.

Do mesmo modo, Shah et al.(2004) utiliza algoritmo genético para atuar na seleção construindo árvores de decisão e Clark et al. (2005; 2008) aplica AG para construir árvores de decisão baseada em expressões booleanas construídas com blocos de SNPs.

Outra combinação de AG foi implementado como uma estratégia adaptativa evolutiva em combinação comum a abordagem baseada em desequilíbrio de ligação para identificar interação de loci (Fontanarosa, Yang; 2010).

* 1. **Abordagens computacionais para GWAS**

Com o conceito de interações epistáticas que são interações não-aditivas entre alelos, *lócus* ou mutações que com efeito combinado entre as mutações não se espera de seus efeitos individuais. Portanto, os genes podem mascarar a presença de outro ou se combinar para produzir uma característica nova (Phillips, 2009). Sendo assim, a variação genética que possui alcance individual suficiente para influenciar sobre uma doença ou fenótipo é conhecida como efeito principal (Yang et al., 2010). Assim, com este cenário de variação genética vem sendo desenvolvido um intenso avanço em ferramentas para pesquisas genéticas, como banco de dados, e tecnologias para genotipagem cada vez mais precisas e acessíveis, contribuindo aos avanços no desenvolvimento de softwares para análise GWAS (Li, *et al.* 2014)

Para Yang (2009), ao se tratar com a análise de dados de GWAS, a imensa quantidade de dados se torna um fator crucial na seleção dos métodos mais adequados, pois prevalece a impossibilidade de todos os métodos serem capazes de lidar com tantos dados em um tempo que não seja proibitivo. O tempo computacional pode ser substancialmente reduzido se uma técnica de seleção de variáveis apropriada é usada. Com base na estratégia de otimização usada, foi definida uma classificação de métodos computacionais: métodos de busca exaustiva, métodos de busca gulosa e métodos de busca estocástica.

Cabe ressaltar que as ferramentas que são aplicadas os métodos serão analisadas neste trabalho gerando resultados comparativos sobre os impactos nas pesquisas.

* + 1. **Método de busca exaustiva**

A busca exaustiva, enumera todos os k-lócus de interações possíveis entre SNPs para identificar o efeito, ou efeitos, que melhor predizem o desfecho fenotípico. Esta propriedade exaustiva leva à sua característica mais importante do ponto de vista computacional. Embora seja viável, mesmo para os maiores conjuntos de dados disponíveis hoje, utilizando, por exemplo, computação paralela, as generalizações para investigar interações de ordem superior são excessivamente demoradas. Nesta seção será realizado um breve resumo de algumas abordagens exaustivas disponíveis na literatura.

* + 1. **Método de busca gulosa**

Os métodos de busca gulosa utilizam de algoritmos que buscam uma solução para o problema tomando decisões que levam a um novo resultado ótimo a cada passo da execução. Com isso, espera-se que ao final da busca seja alcançado o resultado ótimo global sem que seja necessário analisar todas as situações possíveis (Torres e Castro, 2012).

* + 1. **Método de busca estocásticos**

Os métodos estocásticos realizam uma investigação probabilística do espaço de busca. Alguns começam com um modelo composto por um conjunto aleatório de SNPs e tentam melhorar a sua precisão de classificação, enquanto outros usam um subconjunto pequeno que foram previamente selecionados incorporando conhecimento aprimorado sobre os dados (Jiaquiao, et al. 2008).

1. **RESULTADOS E IMPACTOS ESPERADOS**

A presente proposta visa a comparação de ferramentas de Bioinformática, realizando uma análise sistemática dos modelos que possa auxiliar na interpretação aplicada à metodologia de microarranjos, para detecção e predição de CNVs sobre as características de SNPs da população exposta ao RI.

Os resultados desta estratégia de diagnóstico serão importantes para promover um conhecimento sobre os efeitos biológicos de exposição à radiação ionizante.

Ao final do estudo, os resultados permitirão inferir sobre os possíveis mecanismos moleculares associados ao retardo mental idiopático autossômico, além de contribuir para uma melhor elucidação da fisiopatologia dessa alteração genética, tão heterogênea.

Resumidamente, ao final do projeto, são esperadas as seguintes contribuições científicas e biotecnológicas:

1. Encontrar técnicas de comparações em bases de dados distintos baseadas em GWAS;
2. Apresentar vantagens e desvantagens dos métodos de buscas que serão apresentados em estudos;
3. Realizar uma revisão geral sobre as ferramentas que serão utilizadas para investigar efeitos de interação epistática em GWAS;
4. Aplicar técnica de lógica para achados e análise de mutações de novo em pacientes com exposição parental à radiação ionizante de césio-137 a partir de dados de genotipagem de polimorfismos de base única de alta densidade;
5. Definir um motor de *workflow* para auxiliar na compreensão da metodologia desenhada após as conclusões do estudo;
6. Propor os melhores modelos estatísticos em bases de dados;
7. Publicação de artigos científicos que envolvam: descobertas de conhecimentos por meio de algoritmos inteligentes em achados;
8. Estabelecimento e fortalecimento de parcerias com centros de referência e de excelência no diagnóstico molecular das doenças genéticas humanas.
9. **BIBLIOGRAFIA**

|  |  |
| --- | --- |
|  | Betancur C. Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. Brain Research, 1380.42-77, 2011. |
|  | Caetano A. R., Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o future. R. Bras. Zootec., v.38, p.64-71, 2009. |
|  | Coulter ME, Miller DT, Harris DJ, Hawley P, Picker J, Roberts AE, Sobeih MM, Irons M. Chromosomal microarray testing influences medical management. Genet Med. Sep;13(9):770-6, 2011. |
|  | Clark, T.G., de Lorio, M., Griffiths, R.G., Farrall, M. Finding Association in Dense Genetic Maps: A Genetics Algorithm Approach. Human Heredity, v.60, pp. 97–108. 2005. |
|  | Clark, T.G., de Lorio, M., Griffiths, R.G., An Evolutionary Algorithm to Find Associations in Dense Genetic Maps, IEEE Transactions on Evolutionary Computation, v.12, n.3, pp. 297–306. 2008. |
|  | Fombonne E. Epidemiology of autistic disor-der and other pervasive developmental dis-orders. J Clin Psychiatry. V. 66, 3– 8, 2005. |
|  | Fontanarosa J., Yang D., A Block-Based Evolutionary Optimization Strategy to Investigate Gene-Gene Interactions in Genetic Association Studies, Bioinformatics and Biomedicine Workshops (BIBMW), 2010 IEEE International Conference, vol. 330-335. 2010. |
|  | Grafodatskaya D, Chung B, Szatmari P, Weksberg R. Autism Spectrum Disorders and Epigenetics. Jrnlameracad child &adol psych.794-49.8, 2010. |
|  | Goldberg, David E. Genetic Algorithms in Search, Optimization, and Machine Learning. EUA: Addison-Wesley, 1989. |
|  | Greene CS, White BC, Moore JH., Ant Colony Optimization for Genome-Wide Genetic Analysis,Lect Notes Comput Sci., v.5217, pp. 37-47, 2008. |
|  | Kong SW, Shimizu-Motohashi MG, Lee HC, Collins CD, Brewster SJ, Holm IA, Rappaport L, Kohane IS, Kunkel LM. Peripheral blood gene expression signature differentiates children with autism from unaffected siblings. Neurogenetics 14:143–152. 2013. |
|  | Li P., Guo M., Wang C., Liu X., ZouAn O. Overview of SNP interactions in genome-wide association studies. Briefings in functional genomics. Vol 14. no 2. 143-155. 2014. |
|  | Motsinger-Reif AA, Fanelli TJ, Davis AC, Ritchie MD., Power of grammatical evolution neural networks to detect gene - gene interactions in the presence of error, BMC Res Notes. , v. 1, n.65 , pp. 1-8, 2008. |
|  | Nazeer A, Ghaziuddin M. Autism Spectrum Disorders: Clinical Features and Diagnosis. MDbPediatrClin N Am.59,19–25, 2012. |
|  | Phillips P.C., Epistasis the Essential Role of Gene Interactions in the Structure and Evolution of Genetic Systems, 2008, Nature Rev. Genetics, vol. 9, no. 11, pp. 855-867. |
|  | Picker LW, Peter RSJ, Laurie AD, Heather KEB. Caronna, KJ, et al. Clinical Genetic Testing for Patients With Autism Spectrum Disorders. Pediatrics. v.125, no 4, 2010. |
|  | Shah, S., Kusiak, A., Data mining and genetic algorithm based gene/SNP selection, Artificial Intelligence in Medicine, v. 31, pp.183–196, 2004. |
|  | Shi L, Zhang X, Golhar R, Otieno FG, Mingze H, Cuiping H, Kim C, Wang K, Hakonarson K. Whole-genome sequencing in an autism multiplex family. Molecular Autism, 4:8, 2013. |
|  | Smoller J. W., and Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. Vol. 381: 1371–79. 2013 |
|  | The Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. Nature Publishing Group. 2014. |
|  | Vargas EA, Pastore M, Prior T, et al. The preva-lence of PTEN mutations in a clinical pediatric cohort with autism spectrum disorders, developmental delay, and macrocephaly. Genet Med. 11(2): 111–117, 2009. |
|  | Xu Y., Peng B., Yunxin Fu Y., Amos C.: Genome-wide algorithm for detecting CNV associations with diseases. BMC Bioinformatics. 2011, P.12:33. |
|  | Yang, C.,He Z., Wan X., Yang O., Xue H., Yu W.,SNPHarvester: a filtering-based approach for detecting epistatic interactions in genome-wide association studies. Bioinformatics 2009; 25:504-511. |
|  | Yang, C., Wan, X., Yang, Q., Xue, H., Yu, W., 2010, “Identifying main effects and epistatic interactions from large-scale snp data via adaptive group lasso”, BMC Bioinformatics , v.11, Suppl 1, S18. |
|  | Yang, C., Wan, X., Yang, Q., Xue, H., Yu, W., Identifying main effects and epistatic interactions from large-scale snp data via adaptive group lasso”, BMC Bioinformatics, v.11, Suppl 1, S18 |
|  | Yiping S, Kira AD, Ingrid A. Holm MD, Carolyn BMD, et al. Clinical Genetic Testing for Patients With Autism Spectrum Disorders. Pediatrics. v: 125, no 4, 727- 735, 2010. |
|  | Zhang F., Liang H. "Application of Data Mining Techniques in Universal Design," Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2014 International Conference on , vol., no., P.179,182 |
|  | Ziats MN, Rennert OM. Aberrant Expression of Long Noncoding RNAs in Autistic Brain. J MolNeurosci 49:589–593, 2013. |
|  | Frankenberg-Schwager, M. Induction, repair and biological relevance of radiation-induced DNA lesions in eukaryotic cells. Radiat. Environ. Biophys. 29,273–292.1990. |
|  | Nakamura, N. et al. Radiation effects on human heredity. Ann. Rev. Genet 47, 51–68. 2013. |
|  | UNSCEAR. Hereditary Effects of Radiation United Nations (2001). |
|  | ADEWOYE, A. B. et al. The genome-wide effects of ionizing radiation on mutation induction in the mammalian germline. Nature Communications, v. 6, n. 6684, p. 1-8, Mar, 2015. |
|  | da Cruz, A. D. et al. Human micronucleus counts are correlated with age, smoking, and cesium-137 dose in the Goiânia (Brazil) radiological accident. Mutation research, v. 313, n. 1, p. 57-68, Aug, 1994. |
|  | Wang, K.; Li, M.; Hadley, D.; Liu, R.; Glessner, J.; Grant, S.F.; Hakonarson, H.; Bucan, M. PennCNV: Na integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007, 17, 1665–1674. |
|  | Colella, S.; Yau, C.; Taylor, J.M.; Mirza, G.; Butler, H.; Clouston, P.; Bassett, A.S.; Seller, A.; Holmes, C.C.; Ragoussis, J. QuantiSNP: Anobjectivebayeshidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. NucleicAcids Res. 2007, 35, 2013–2025. |
|  | Korn, J.M.; Kuruvilla, F.G.; McCarroll, S.A.; Wysoker, A.; Nemesh, J.; Cawley, S.; Hubbell, E.; Veitch, J.; Collins, P.J.; Darvishi, K.; et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat. Genet. 2008, 40, 1253–1260. |
|  | Marioni, J.C.; Thorne, N.P.; Valsesia, A.; Fitzgerald, T.; Redon, R.; Fiegler, H.; Andrews, T.D.; Stranger, B.E.; Lynch, A.G.; Dermitzakis, E.T.; et al. Breaking the waves: Improved detection of copy number variation from microarray-based comparative genomic hybridization. Genome Biol. 2007, 8, R228. |
|  | Diskin, S.J.; Li, M.; Hou, C.; Yang, S.; Glessner, J.; Hakonarson, H.; Bucan, M.; Maris, J.M.; Wang, K. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. 2008, 36, e126. |
|  | S.P.Torres andC.A.Castro, Parallel particles warm optimization applied to the static Transmission Expansion Planning problem,Transmissionand Distribution: Latin America Conference and Exposition (T&D-LA), 2012 Sixth IEEE/PES, Montevideo, 2012, pp. 1-6. doi: 10.1109/TDC-LA.2012.6319053 |
|  | Hu, Jiaqiao; Fu, Michael C.; Marcus, Steven I. A Model  Reference Adaptive Search Method for Stochastic Global Optimization. Commun. Inf. Syst. 8 (2008), no. 3, 245--276. http://projecteuclid.org/euclid.cis/1241018503. |
|  |  |