Updated Cluster Generation Protocol for X Chromosome SNPs

An improved cluster generation protocol that uses only female samples for clustering X chromosome SNPs provides higher quality data and better signal-to-noise ratios.

Overview

For the Infinium® HD product line (including Human610-Quad and Human1M-Duo BeadChips), the cluster file (*.egt) supplied by Illumina has been generated with an updated process to improve copy number data quality on the X chromosome.

The cluster file was generated by clustering the X chromosome SNPs using female HapMap samples only, rather than a mixture of males and females. This was done to reduce the noise in the X chromosome data from the males and to provide a more representative reference for log R ratio calculations in customer samples. This Technical Note describes how to implement this clustering strategy with any sample data set.

Clustering Protocol

Use the following procedure to create a cluster file in which the X chromosome loci are clustered using only the female samples.

1. **Load a BeadStudio project workspace (*.bsc file) containing all of the sample data.**
2. **Select all of the male samples in the Samples Table.**
3. **Right-click the selected samples to launch the context menu.**
4. **Select Exclude Selected Samples. This ensures that only female samples are used by the clustering algorithm.**
5. **Filter the SNP Table so that only X chromosome loci with genotype calls are displayed in the SNP table. If there are any intensity only X loci, filter out these loci as well. Intensity only loci are identified by a 1 in the Intensity Only column.**
6. **Select all displayed X chromosome loci.**
7. **Right-click to launch the context menu and select Recluster Selected Loci. Do not update statistics at this point, because males need to be reintroduced into the project first.**
8. **Select all of the males previously excluded in the samples table.**
9. **Right-click to launch the context menu and select Include Selected Samples.**
10. **In the SNP table, select the displayed X loci.**
11. **Right-click and select Update Selected SNP Statistics.**
12. **To properly include these changes in any report that may be generated (such as the DNA report), select Analysis | Update Heritability and Reproducibility Errors and update DNA statistics by clicking the calculate button ( ) in the Samples Table.**

Conclusion

After following this protocol, all parameters, including the log R ratio, will be calculated to reflect this updated clustering scheme. Incorporating the highest quality data and consistency with Illumina's supplied cluster file will ensure high performance of downstream CNV analysis.

Additional Information

For more information about cluster generation and Illumina DNA Analysis BeadChips, please contact us at the address below.

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2010 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, GenomeStudio, Genetic Energy, HiSeq, and HiScan are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 370-2008-009 Current as of 21 March 2008

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA