Quality Process Control with Geneticist Assistant

July 2014

Megan McCluskey, Edward Bouton

Introduction

Geneticist Assistant NGS Interpretive Workbench is designed for the management of next generation sequencing data, allowing the creation of a database of analysis information including variant details, including confirmed pathogenicity calls, coverage information, sample and patient information, and historical variant information. Additionally, Geneticist Assistant reports coverage and variant frequencies, including positive controls, over time that can be useful to track trends in test performance.

Coverage of all regions within a panel is reported for each sample, as well as average values across all samples that have been added to the database. A date range can be adjusted to report coverage values across different time periods. This can be helpful for identifying any changes in performance that can then be quickly addressed.

Panel DLMP Panel Group default for samples added between 5/1/2014 and 5/31/2014 Refresh

Coverage Regions of Panel 'DLMP':

ID	Region Name	Chrom:Start - End	Average % Covered	Average Average Coverage	Average Minimum Coverage	Passed	Passed Percent	Failed	Total	Gene
154	CHEK2:NM_007194	22: 29130361 - 29130739	100%	138	90	10	90.9091%	1	11	CHEK2
153	CHEK2:NM_007194	22: 29121201 - 29121385	100%	163	110	10	90.9091%	1	11	CHEK2
152	CHEK2:NM_007194	22: 29120935 - 29121142	100%	197	84	10	90.9091%	1	11	CHEK2
151	CHEK2:NM_007194	22: 29115353 - 29115503	100%	325	277	10	90.9091%	1	11	CHEK2
150	CHEK2:NM_007194	22: 29107867 - 29108035	100%	460	321	10	90.9091%	1	11	CHEK2
149	CHEK2:NM_007194	22: 29105964 - 29106077	100%	79	61	2	18.1818%	9	11	CHEK2
148	CHEK2:NM_007194	22: 29099463 - 29099584	100%	318	263	10	90.9091%	1	11	CHEK2
147	CHEK2:NM_007194	22: 29095796 - 29095955	100%	311	247	10	90.9091%	1	11	CHEK2
146	CHEK2:NM_007194	22: 29092859 - 29093005	100%	389	275	10	90.9091%	1	11	CHEK2
145	CHEK2:NM_007194	22: 29091668 - 29091891	100%	301	134	10	90.9091%	1	11	CHEK2
144	CHEK2:NM_007194	22: 29091085 - 29091260	100%	553	452	10	90.9091%	1	11	CHEK2
143	CHEK2:NM_007194	22: 29089990 - 29090135	100%	480	378	10	90.9091%	1	11	CHEK2
142	CHEK2:NM_007194	22: 29085093 - 29085233	100%	161	125	10	90.9091%	1	11	CHEK2
141	CHEK2:NM_007194	22: 29083855 - 29084004	100%	366	152	10	90.9091%	1	11	CHEK2
140	STK11:NM_000455	19: 1226423 - 1226676	100%	493	239	10	90.9091%	1	11	STK11
139	STK11:NM_000455	19: 1222954 - 1223201	100%	497	259	10	90.9091%	1	11	STK11
138	STK11:NM_000455	19: 1221918 - 1222035	100%	198	133	10	90.9091%	1	11	STK11
137	STK11:NM_000455	19: 1221182 - 1221369	100%	596	391	10	90.9091%	1	11	STK11
136	STK11:NM_000455	19: 1220550 - 1220746	100%	559	315	10	90.9091%	1	11	STK11
135	STK11:NM_000455	19: 1220342 - 1220534	100%	606	396	10	90.9091%	1	11	STK11

Figure 1: The Coverage Regions table showing average coverage values based on samples submitted in May 2014.

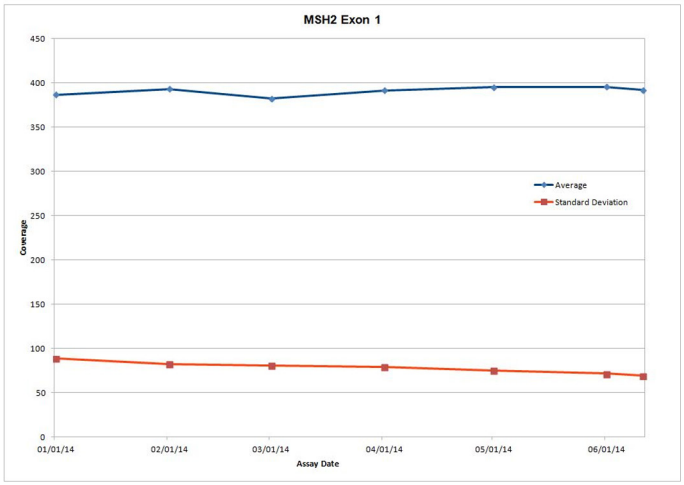

Figure 2: Average coverage and standard deviation for exon 1 in the MSH2 gene for monthly time points from January through June 2014

For quality control, repeating analysis for the same sample to ensure consistent results is a valuable tool. Geneticist Assistant's reporting can facilitate this process. Sample groups, for example a control group, can be created to track variant statistics relative to these specific samples that may be sequenced repeatedly over time. Utilizing samples in the designated group, minimum, maximum, mean, median and standard deviation values for coverage, variant frequency and read balance are reported for a variant. Again, a time range setting is available to allow reporting statistics for different time periods.

ID~	Chr : ChrPos	Gene	Variant Frequency	Coverage	HGVS Coding	Control: *06.17.2014-07.23.2014 Number of Samples	Control: Number of Samples having the variant	Control: Coverage Min	Control: Coverage Max	Control: Coverage Median	Control: Coverage Mean	Control: Coverage Standard Deviation	Control: Variant Frequency Min	Control: Variant Frequency Max	Control: Variant Frequency Median	Control: Variant Frequency Mean	Control: Variant Frequency Standard Deviation
6	3:37083740	MLH1	0.5	416	c.1668-19A>G	15	12	220.0	416.0	327.0	315.0	69.1483911599	0.5	1.0	0.5	0.541666666667	0.138192699598
12	14:75483812	MLH3	1	514	c.4335A>G	15	11	274.0	514.0	404.0	396.181818182	82.7853941683	0.5	1.0	1.0	0.818181818182	0.24052284646
13	14:75505016	MLH3	1	564	c.3643+27T>C	15	15	301.0	564.0	417.0	408.866666667	83.908574585	1.0	1.0	1.0	1.0	0.0
14	14:75513883	MLH3	1	501	c.2476A>G	15	15	294.0	501.0	381.0	381.8	68.0001960781	1.0	1.0	1.0	1.0	0.0
15	17:7579472	TP53	1	580	c.215C>G	15	12	349.0	580.0	458.0	485.75	65.9534242225	0.5	1.0	0.75	0.75	0.25
18	17:63554591	AXL...	0.5	950	c.148C>T	15	10	522.0	950.0	685.0	700.8	146.279048397	0.5	1.0	1.0	0.8	0.244948974278
25	2:48032875	MSH6	1	384	c.3646+39_36...	15	9	205.0	453.0	384.0	336.111111111	92.894736658	0.5	1.0	0.5	0.611111111111	0.207869854821
41	2:48023115	MSH6	0.5	991	c.5407>C	15	5	615.0	991.0	807.0	842.2	140.271736284	0.5	0.5	0.5	0.5	0.0
42	2:48025764	MSH6	0.5	442	c.642C>T	15	4	442.0	491.0	466.5	466.5	24.5	0.5	0.5	0.5	0.5	0.0
45	2:48010488	MSH6	0.5	614	c.116G>A	15	3	412.0	614.0	614.0	546.666666667	95.2237131998	0.5	0.5	0.5	0.5	0.0

Figure 3: Sample Group Statistics showing the minimum, maximum, mean, median and standard deviation values for coverage and variant frequency

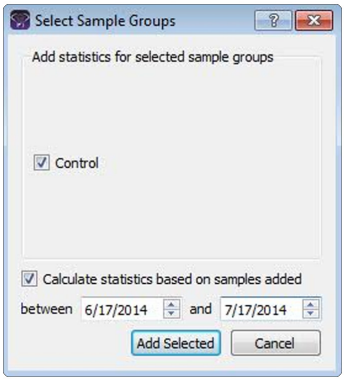

Figure 4: Sample Group Statistics can be managed to report values based on samples added during a user-defined time period. Changes in assay performance can be detected by tracking statistics over different time periods.

Variant frequency values for control samples can be exported to Excel to create graphs for a variety of statistical relationships, for instance to check the relationship between the coefficient of variation and the allele frequency.

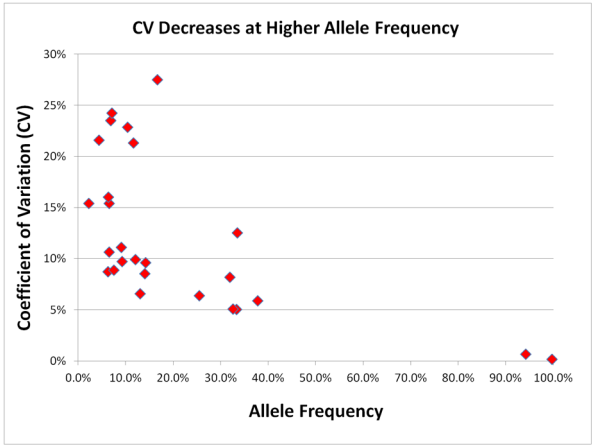

Figure 5: Coefficient of Variation Decreases at Higher Allele Frequency. Sample group variant frequency statistics were exported from Geneticist Assistant to determine the relationship between allele frequency and assay variation. The Coefficient of Variation (y-axis) is plotted against the variant allele frequency (x-axis) from 6 separate assay runs of a control material. The control material has multiple variant allele frequencies created from admixture of cancer cell lines. The %CV decreases with higher allele frequencies. The %CV is less than 30% for all variant frequencies in all runs of the control material. The assay has a %CV of less than 20% when the allele frequency is greater than 20%. This information can be used to establish the limit of detection of the assay. Data courtesy of Children's Medical Center – Dallas.