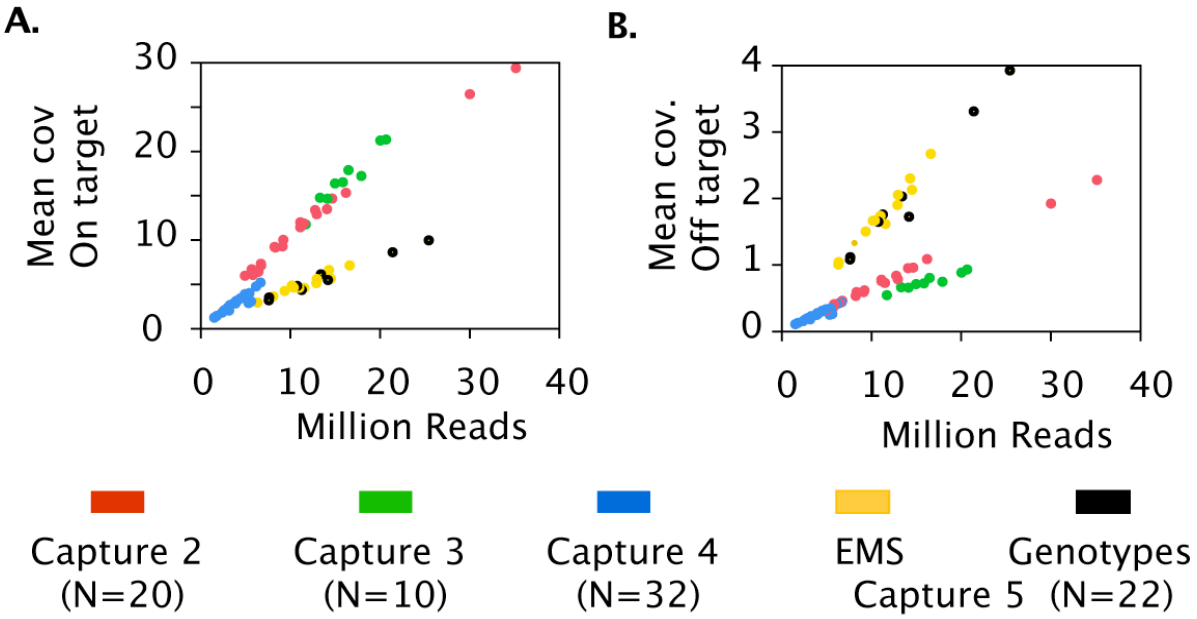
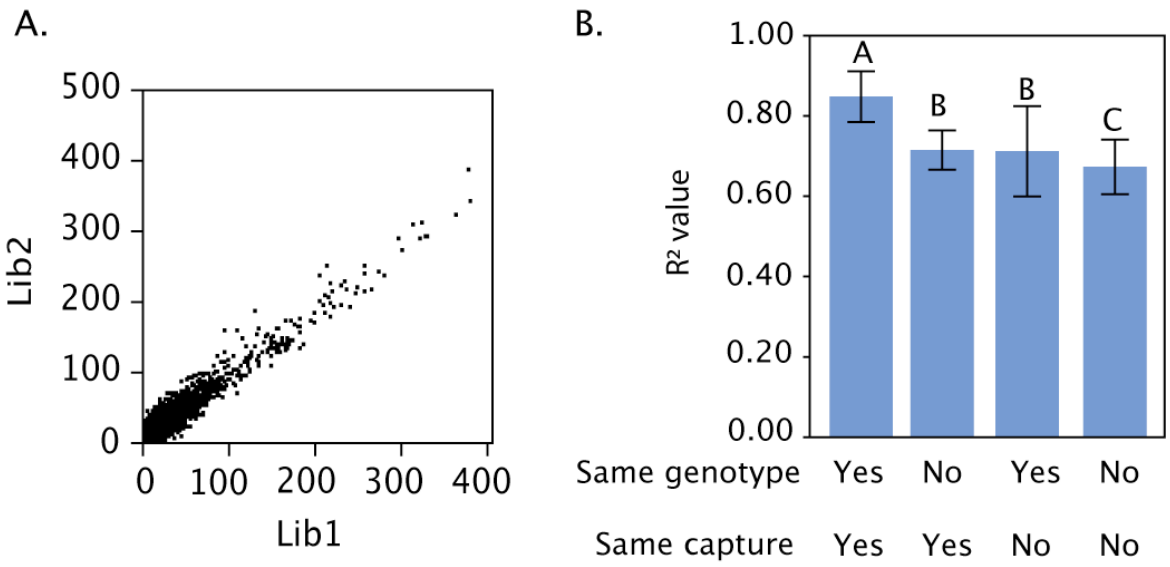
Supplemental Figure 1: Mean coverage statistics for targeted and non-targeted regions of the rice genome, for each sample.

For each sample, mean coverage per bp was calculated for all positions included in the capture targets (A) and for all other positions (B). The relationship between coverage and number of sequencing reads obtained for each library is shown. Samples processed in the same capture reaction are labelled in the same color, with the exception of the samples present in capture 5, which are divided into "EMS samples" and "genotypes". * Capture #1 failed at the sequencing level and is therefore not included in this figure.



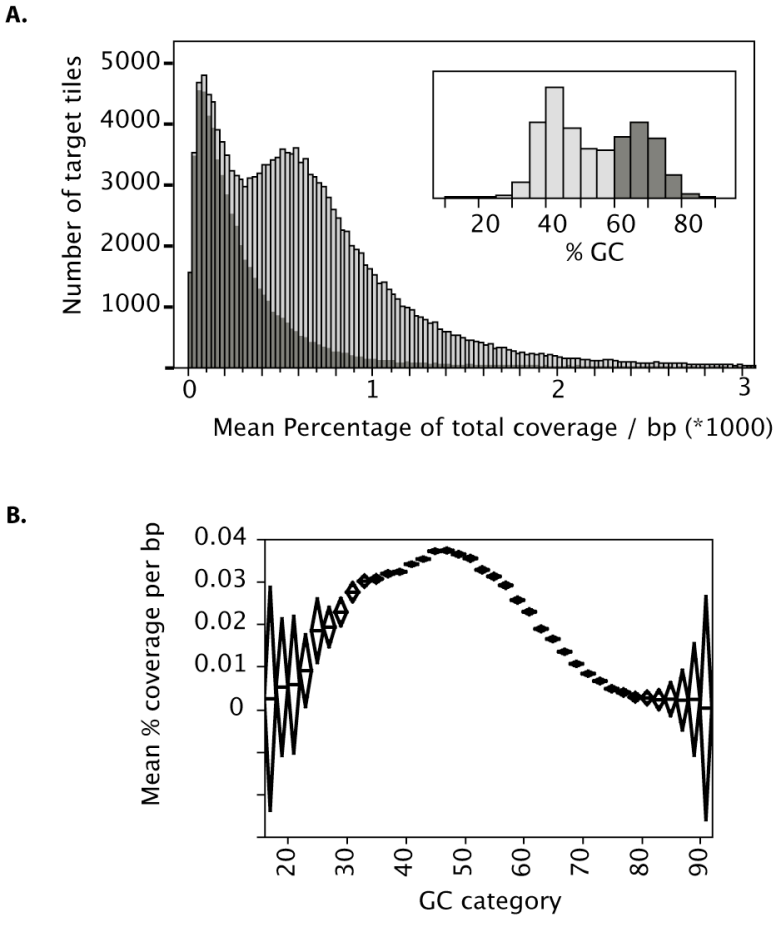
Supplemental Figure 2 - Consistency of capture between rice samples.

For each sample, the number of reads mapping to each target tile was calculated. Next, these coverage numbers were compared between samples on a pair-wise basis. **A.** Example of the correlation between tile coverage in two sample processed in the same capture. Each dot represents a target tile. The regression p-value and R^2 values are indicated. **B.** Comparisons were divided into four categories, based on whether the two samples were of the same genotype or not and whether they were part of the same capture reaction or not. For each category, mean R^2 and standard deviation values were calculated.



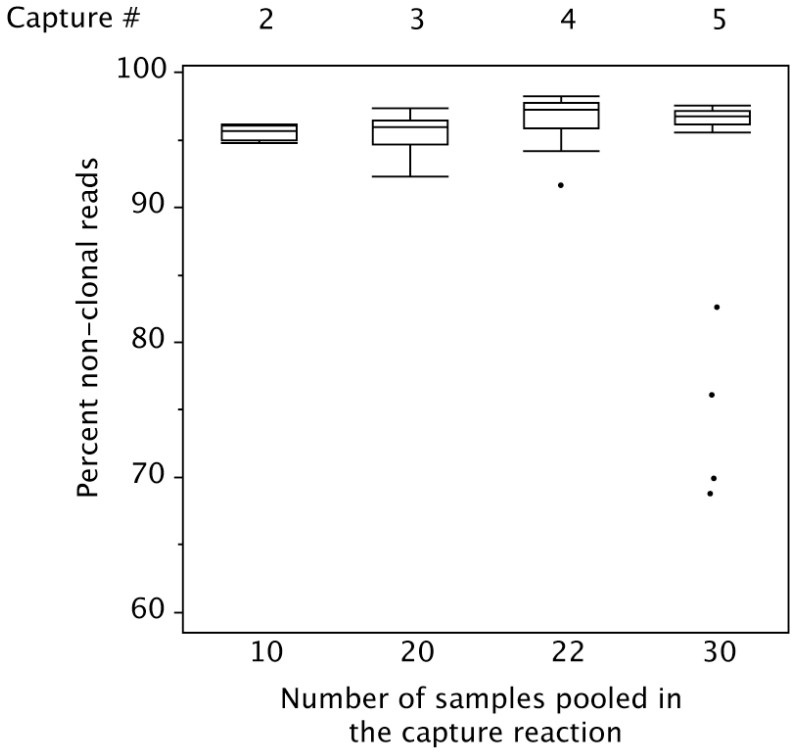
Supplemental Figure 3 - Effect of GC content on rice target coverage.

A. Target coverage was expressed as the percentage of reads mapping to each target tile, and normalized based on target length. For each tile, the mean target coverage was calculated and the distribution of values is represented. Inset: distribution of GC content per target tile. Selection of tiles with a GC content above 60% corresponds to a highly biased subset of target tiles associated with low target coverage (dark gray subset on both graphs). **B.** Direct effect of target tile GC content on the number of reads captured. Target tiles were divided into categories based on GC content (5% categories). For each GC category, the mean is indicated by a straight line and the standard deviations are indicated by diamond shapes centered on the mean values. The disctribution of target coverage were discarded before calculating the mean percentage of coverage per genome content category.



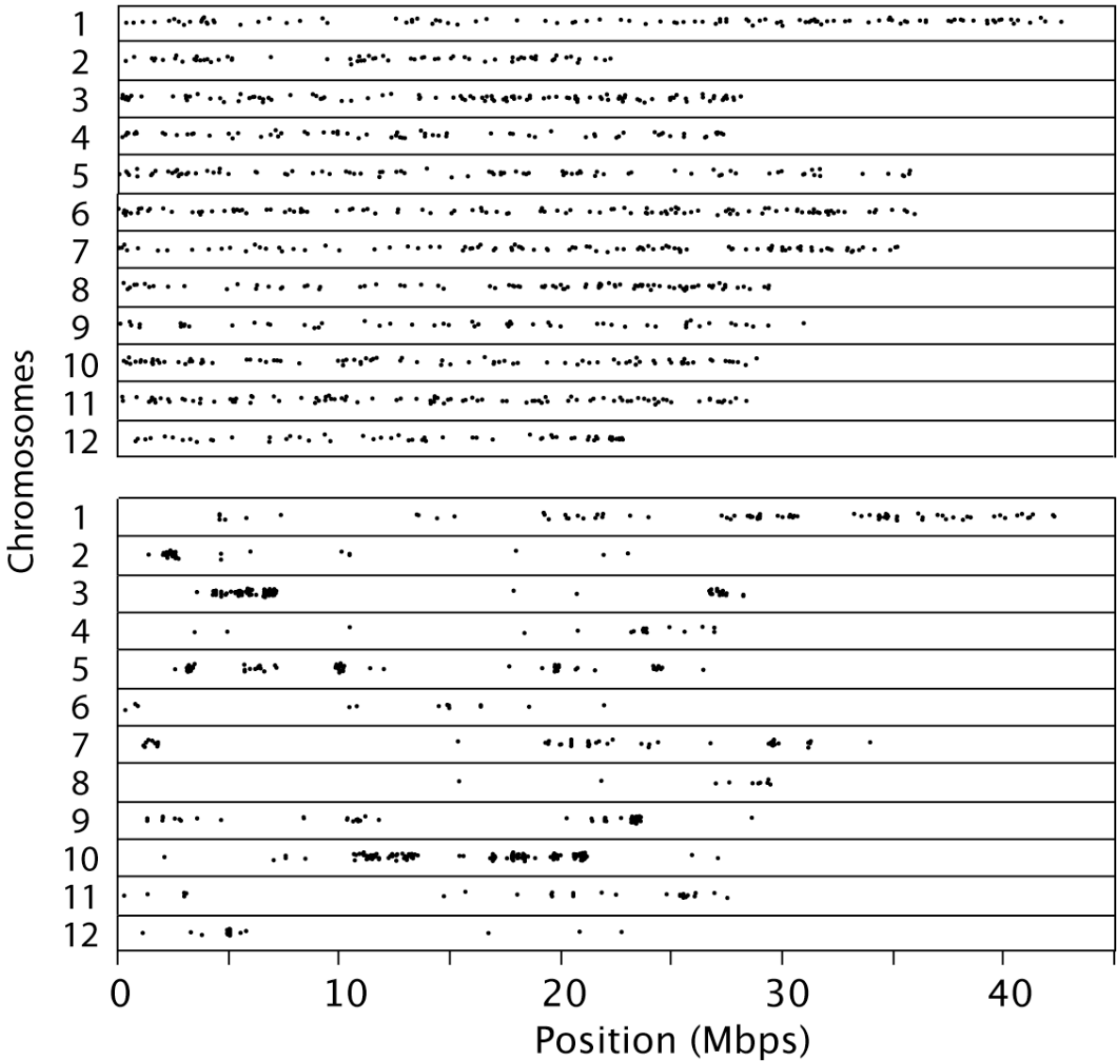
Supplemental Figure 4 - Effect of sample pooling prior to sequence capture on the presence of clonal reads.

Up to 28 rice genomic libraries were pooled together prior to performing the capture reactions. It is possible that a high level of pooling could be detrimental to downstream analyses. Indeed, it is possible that, if too few fragments are contributed from each library, the resulting captured reads will have low complexity. This can be assessed by measuring the percentage of unique reads in each sample. To be able to compare samples to each other, each read file must contain the same number of reads. Therefore, 2.5 million reads were randomly selected from each of the samples for which at least 2.5 million reads had been obtained. Next, the percentage of unique reads (reads that have a unique starting position and direction after mapping to the reference genome) was calculated (see "Read processing" in the Methods section). Observing a lower percentage of unique reads from samples that experienced a higher level of pooling would be diagnostic of a detrimental effect of pooling. For each capture, the mean percentage of unique reads was calculated. The means and standard deviations are shown. Outliers are indicated by dots. There were no significant differences between the means observed for the different captures. Capture #1 failed at the sequencing level and is therefore not included in this figure.



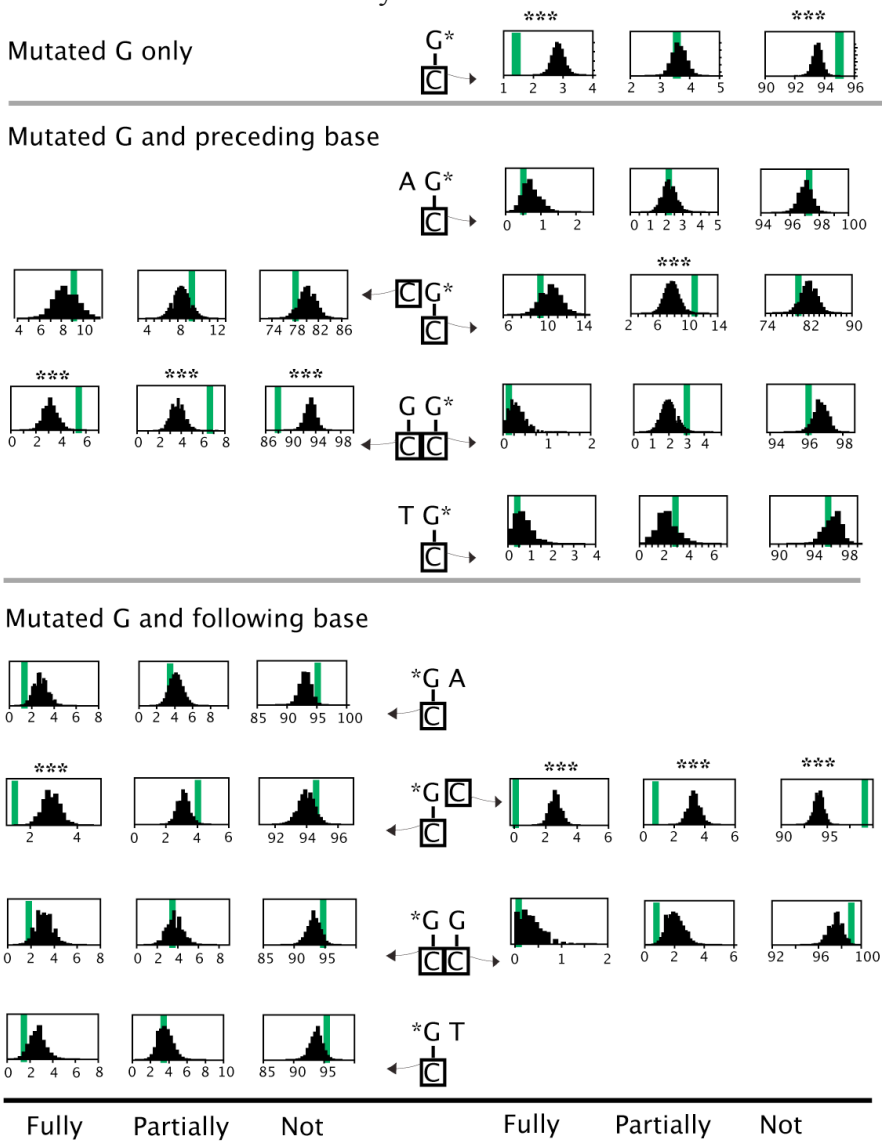
Supplemental Figure 5 - Distribution of mutations along the 12 rice chromosomes in an EMS-mutagenized sample and a potential seed contaminant.

One of the rice EMS-mutagenized samples exhibited a lower percentage of expected (CG>TA) mutations than all other samples. To test whether this sample is a seed contaminant, the location of the mutations found in this sample were plotted along the twelve chromosomes of the rice genome (B). A control individual is shown on top for comparison (A). In the control sample, mutations are evenly distributed along the chromosomes while the potential seed contaminant exhibits islands of high mutation density and regions of poor mutation density, reminiscent of potential introgressed regions from a different genotype.



Supplemental Figure 6 – Relationship between cytosine methylation and EMS targeting depending on sequence context in rice.

Observed (thick green vertical lines) and expected (distribution of values) percentages of fully methylated (Fully), partially methylated (Partially) and unmethylated (Not) cytosines opposite and / or flanking the mutated guanines. For each graph, the thick green line represents the observed percentages from the mutated positions. The number of positions included in the calculation of those percentages depends on the number of mutations (N) for which methylation data were available. The distribution of expected percentages upon random selection of N nucleotides or dinucleotides for which methylation data are available is shown in black (100,000 random samplings). The top panel shows data for all mutated guanines at once. The bottom two panels depict how these percentages vary depending on the nucleotide context. G*: guanine residues that were found to be mutagenized in our captured individuals. The cytosine residue for which the methylation state is evaluated is surrounded by a black square. ***: less than 10/100,000 random samples exhibited values further from the mean of the distribution than the observed mean (green line).



Supplemental Table 1 - Parameters used for mutation detection using the MAPS bioinformatics pipeline for each of the capture reactions.

All other parameters were set to the default values.

Capture number*	2	3	4	5	Wheat
Number of libraries	20	10	28	22	8
Minimum # libraries covered (-l)	5	4	7	5	5
Minimum total coverage (-v)	20	20	20	20	10
MAPS 1-specific parameters					
Maximum coverage (-c)	10,000	5,000	15,000	11,000	5,000
Min. % of each het. allele (-i)	5	5	5	5	5
MAPS 2-specific parameters					
Min.% of mutant het. allele (-d)	20	20	20	20	15
Min. cov. of mutant het. allele (-p)	4	4	4	4	5-7
Min. cov. of mutant homoz. allele (-s)	3	3	3	3	4-5

* Capture #1 failed at the sequencing level and is therefore not included in this report. Het. = Heterozygous. Homoz. = Homozygous

Supplemental Table 2 – List of mutations selected for PCR validation.

Chr.	Pos.	Conf.	Tot. Cov.	Lib.	Capture	Ho/He	WT Cov.	MA Cov.	Type	Lib. Cov.	# libs	Gene Model
chr1	1933530	No	137	uN5	2	hom	0	5	GA	5	19	LOC_Os01g04340.1
chr1	21118556	No	60	uN8	2	hom	0	3	GA	3	15	LOC_Os01g37760.1
chr4	17856816	No	497	uN18	2	het	21	12	CT	33	20	LOC_Os04g30210.1
chr1	8084712	Yes	299	uN2	2	hom	0	6	CT	6	19	LOC_Os01g14440.1
chr1	28937946	Yes	383	uN7	2	hom	0	21	GA	21	20	LOC_Os01g50410.1
chr1	32082308	Yes	36	uE1725	3	hom	0	5	CT	5	10	LOC_Os01g55710.1
chr1	34481567	Yes	190	uE2067	3	hom	0	16	CT	16	10	LOC_Os01g59620.1
chr10	1905155	Yes	184	uN17	2	hom	0	6	TA	6	19	LOC_Os10g04120.1
chr10	17966415	Yes	251	uE2093	3	hom	0	12	GA	12	10	LOC_Os10g33930.1
chr11	5088590	Yes	105	uE2052	3	hom	0	9	GA	9	10	LOC_Os11g09478.1
chr6	15968921	Yes	370	uN17	2	hom	0	9	GA	9	20	LOC_Os06g28124.1
chr2	33927444	Yes	43	uE2093	3	hom	0	4	GA	4	10	LOC_Os02g55400.1
chr11	16970380	Yes	172	uE1719	3	het	8	11	CA	19	9	LOC_Os11g29990.1
chr11	21851803	Yes	194	uE1725	3	het	14	11	GA	25	10	LOC_Os11g37740.1
chr12	9475488	Yes	215	uN20	2	het	21	17	CA	38	20	LOC_Os12g16540.1
chr12	10018408	Yes	258	uN20	2	het	13	13	GA	26	20	LOC_Os12g17490.1
chr5	26081241	Yes	98	uE1725	3	het	4	9	GA	13	10	LOC_Os05g44970.1
chr6	21106121	Yes	283	uE1733	3	het	18	19	GA	37	9	LOC_Os06g36080.1
chr8	17344642	Yes	322	uE1733	3	het	13	18	GA	31	10	LOC_Os08g28410.1
chr8	25102686	Yes	275	uE1733	3	het	15	7	GA	22	10	LOC_Os08g39640.1
chr9	6116830	Yes	168	uE2052	3	het	7	10	AT	17	10	LOC_Os09g11020.1
chr9	17093064	Yes	221	uE1733	3	het	12	8	GA	20	10	LOC_Os09g28180.1

Chr.: Chromosome, **Pos.:** Position, **Conf.:** Sanger sequencing confirmed the mutation (yes/no), **Tot. Cov.:** Total coverage at that position for all samples in that capture experiment, **Lib.:** Name of the sample exhibiting the mutation, **Capture**: Capture experiment. Capture #1 failed at the sequencing level and is therefore not included in this report. **Ho/He:** The mutation was predicted to be homozygous or heterozygous, **WT Cov.:** Coverage of the WT allele for the sample carrying the mutation, **MA Cov.:** Coverage of the mutant allele for the sample carrying the mutation, **Type:** Type of mutation observed (WTallele-Mutant allele), **Lib. Cov.:** Total coverage for the sample carrying the mutation, **# libs:** Number of libraries exhibiting a coverage of at least one at that position, in that capture reaction, **Gene Model:** Name of the gene model for which the mutation is predicted to result in a mis-sense mutation