DESIGN REPORT

Methyl_UniversityofBristol_lung-cancer-risk-panel_1X_MTE-93452736_hg19_LOW

4th September 2023

Lionel Waschnig
Application Scientist, Bioinformatics

@TwistBioscience #SyntheticDNA

Methyl_UniversityofBristol_lung-cancer-risk-panel_1X_MTE-93452736_hg19_LOW

Data received by Twist Bioscience

3904 genomic coordinates in hg19.

Methyl_UniversityofBristol_lung-cancer-risk-panel_1X_MTE-93452736_hg19_LOW

Probe shifting was allowed with an extension of up to 100bp, in addition to usage of a low stringency repeat filter to maximize coverage. Note that this may increase off target rate.
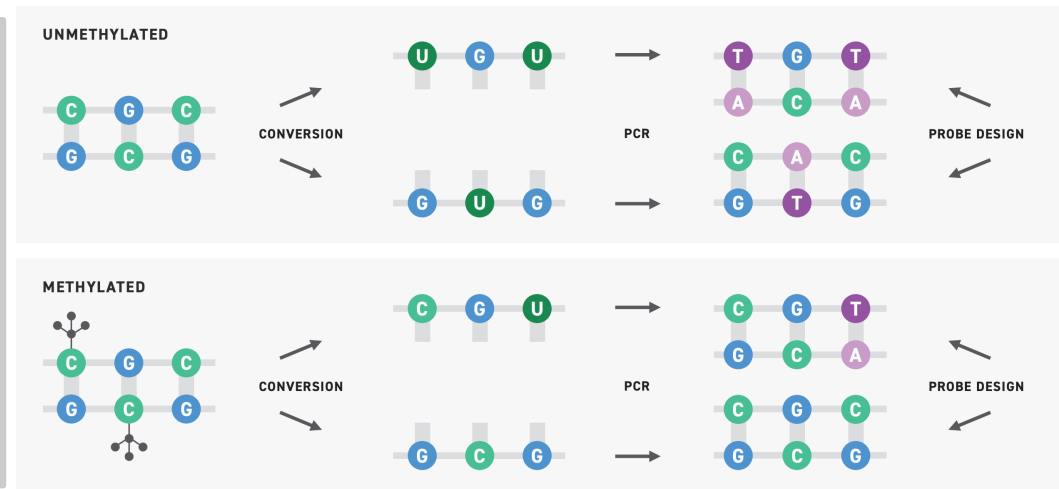
Pre-capture conversion
- Highly complex design algorithm

Four Probe Species Synthesized
- Sense and Anti-sense strands
- Fully methylated and unmethylated

Machine learning strategy
- Imperial data analysis for adaptive design
- Integrated stringency filtering

Methyl_UniversityofBristol_lung-cancer-risk-panel_1X_MTE-93452736_hg19_LOW

Design Summary

Number of probes: 4005

Coverage: 95.86%

Methyl_UniversityofBristol_lung-cancer-risk-panel_1X_MTE-93452736_hg19_LOW

Design Information	
Customer	na
Design Name	Methyl_UniversityofBristol_lung-cancer-risk-panel_shift100_1X_MTE-93452736_hg19_LOWStringency
DesignID	na
Genome	hg19
Probe Length	120
Coverage Summary	
Target Size (bp)	33091
Total Merged Target Regions	3904
Target Merged Regions with Probes	3891
Number of probes	4005
Number of probes removed due to repeats	28
Non Covered Whole Target Regions (#)	24
Non Covered Whole Target Regions (%)	4.14
Non Covered Whole Target Regions (bp)	1370
Covered Size (bp)	31721
Design Size (bp)	475760
Overall Coverage (%)	95.86
Design Files	
all_target_segments	target regions submitted for probe design
all_target_segments_covered_by_probes	all the target regions that are covered at 100% by probes
all_target_segments_not_covered_by_probes	all the target regions that are not covered by probes
merged_probe_file	all the genomic regions that are covered by probes
UCSC_Custom_Tracks	combined covered, not covered, and targets in a UCSC format with header

Methyl_UniversityofBristol_lung-cancer-risk-panel_1X_MTE-93452736_hg19_LOW

Files shared

UCSC_Custom_Tracks_Methyl_UniversityofBristol_lung-cancer-risk-panel_shift100_1X_MTE-93452736_hg19_LOWStringency_230904133216.bed

all_target_segments_Methyl_UniversityofBristol_lung-cancer-risk-panel_shift100_1X_MTE-93452736_hg19_LOWStringency_230904133216.bed

all_target_segments_covered_by_probes_Methyl_UniversityofBristol_lung-cancer-risk-panel_shift100_1X_MTE-93452736_hg19_LOWStringency_230904133216.bed

all_target_segments_not_covered_by_probes_Methyl_UniversityofBristol_lung-cancer-risk-panel_shift100_1X_MTE-93452736_hg19_LOWStringency_230904133216.bed

merged_probe_file_shareable_Methyl_UniversityofBristol_lung-cancer-risk-panel_shift100_1X_MTE-93452736_hg19_LOWStringency_230904133216.bed

report_Methyl_UniversityofBristol_lung-cancer-risk-panel_shift100_1X_MTE-93452736_hg19_LOWStringency_230904133216.html

Definitions

- Target_bases_covered_by_probes: all target regions that are covered at 100% by probes

- Target_bases_not_covered_by_probes : target regions that are not completely covered by probes

- report: Details number of 1) input targets 2) targets covered and 3) number of probes

- Target_regions_with_zero_probes: the target regions for which not a SINGLE probe could be placed for the WHOLE target region.

- All_tracks: combined covered, not covered, whole not covered, and targets in a UCSC format with header

- Probes_merged_ok : all the genomic regions that are covered by probes

# Appendix

Evaluating your custom panel

The bed files provided can be used to evaluate the following:

- Your target regions covered by probes:
  Target_bases_covered_by_probes.bed
- The merged probe coordinates: **Probes_merged_ok.bed**
- Your target regions which are NOT covered by probes:
  Target_bases_not_covered_by_probes.bed /
  Target_regions_with_zero_probes.bed

Consult the **Target_bases_not_covered_by_probes.bed** to assess if any of these uncovered regions are critical for your panel. If targets essential to your panel are filtered out because they occur in repeat regions, Twist can reintroduce probes. Doing so, however, will **increase off-target rates** and **lower the performance of your panel**. In addition, these regions may be challenging to uniquely map in downstream analysis and may result in poor coverage.

The bed files can be loaded into a text editor or into visualiser such as UCSC's genome browser or Integrative Genomics Viewer (IGV).

Visualizing your panel in IGV
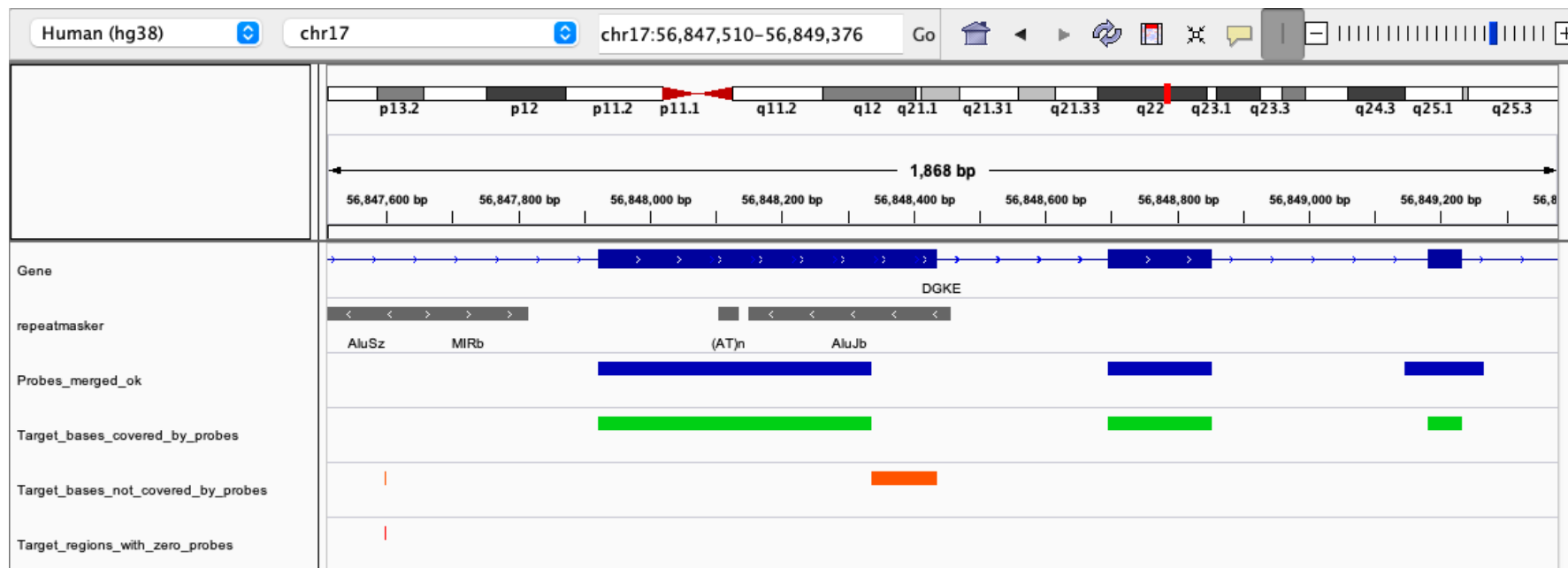
Your bed files can be loaded into IGV for quick and easy visualisation of your covered and non covered target regions.

Track 1: Repetitive elements from repeat masker. Probes that overlap significantly with repeats are filtered out to limit off target rate.

**Track 2: Probes_merged_ok,** shows the merged probe footprint of your panel.

Track 3: Target_bases_covered_by_probes, are the target regions directly covered by probes.

Track 4: Target_bases_not_covered_by_probes, shows the bases within your target regions which are uncovered by probes. This can include partial as well as whole target regions which have been filtered due to repetitive elements.

Track 5: Target_regions_with_zero_probes shows the target regions which are completely uncovered by probes due to the presence of repetitive elements. No probes exist anywhere else on the target coordinate. These targets have been completely filtered out of your panel.
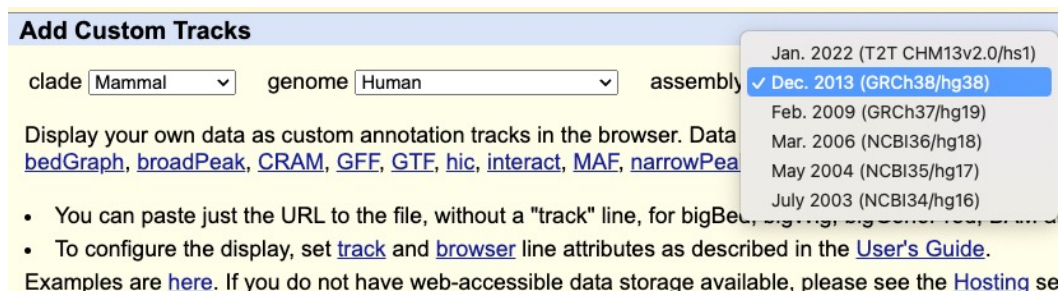
Visualizing your panel in UCSC Genome Browser

1. Navigate to: https://genome.ucsc.edu/
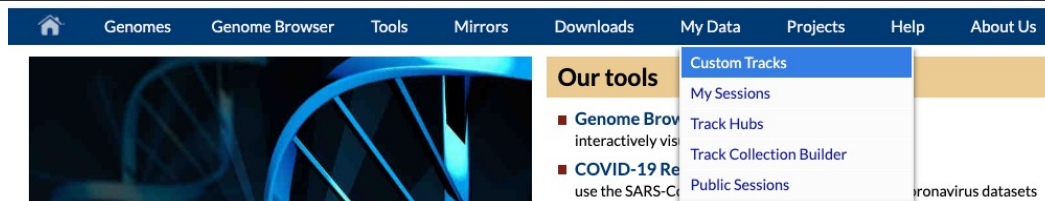
2. My Data > Custom Tracks

3. Under Add Custom Tracks select your genome and assembly.

4. Upload the All_tracks.bed file and submit.

5. View in Genome Browser

https://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html

Visualizing your panel in UCSC Genome Browser

Once in the Genome Browser, your bed file will be represented as colored tracks:

Track 1: Repetitive elements from **RepeatMasker.** Probes that overlap significantly with repeats are filtered out to limit off target rate.
Track 2: Targets, shows all original submitted targets for your panel.
Track 3: Covered, are the target regions directly covered by probes.
Track 4: NOT_Covered, shows the bases within your target regions which are uncovered by probes. This can include partial as well as whole target regions which have been filtered due to repetitive elements.
Track 5: NOT_Covered_Whole shows the target regions which are completely uncovered by probes due to the presence of repetitive elements. No probes exist anywhere else on the target coordinate. These targets have been completely filtered out of your panel.

https://genome.ucsc.edu/goldenPath/help/hgTracksHelp.html