

CGD - Ensembl

- Created using Enseml Perl API
- Selected all exons of all transcripts (all biotypes) for CGD genes based on HGNC gene symbol

CGD - Ensembl flanked 20

• Added 20 bp flanking sequence

CGD - ASD

- Created using online Agilent SureDesign tool
- Selected all exons of all transcripts for CGD genes based on HGNC gene symbol (exceptions: ...)

CGD - ASD flanked 20

- Added -21 bp +20 bp flanking sequence, because ASD erroneously has all start positions +1
- CGD merged Ensembl + ASD flanked with 20bp

Agilent SureSelect All Exon - padded

- Derived from original *_Padded.bed: sorted and removed *chr* prefix from chromosome names
- Original *_Padded.bed = original *_Coverage.bed flanked with 100 bp
- Original *_Coverage.bed contains regions of probes for regions of interest (targets) +
- Original *_Coverage.bed contains by-catch when probe sequence is not unique due to pseudogenes etc.

Agilent SureSelect All Exon - fragment targets

- Derived from original *_fragment_targets.bed: sorted and removed chr prefix from chromosome names
- Contains regions of interest used for probe design
- A subset of the original *_Padded.bed, but NOT a subset of the original *_Covered.bed

Agilent SureSelect All Exon - fragment targets flanked 20

- Added 20 bp flanking sequence
- NOT a subset of the original *_Padded.bed

Agilent SureSelect All Exon - merged: fragment targets flanked 20 & Padded

- Padded regions may be flanked with 0 up to max. 20 bp flanking sequence
- NOT a subset of the original *_Padded.bed

With CGD from 2014-03-12: **3 genes completely missing** in Ensembl75: ATXN8, FSHMD1A and KCNJ18.

With CGD from 2014-03-12: **29 genes completely missing** in ASD: ATXN8, FSHMD1A, KCNJ18 and all 26 MT-* genes. Note KCNJ18 appears present, but is erroneously mapped to region of KCNJ12

WARNING: probe regions may be shifted compared to targets! This allows the capturing kit to compensate for high | low %GC, high similarity to other non-target regions, etc.

Shifting probe regions results in less coverage bias introduced by the kit.

Agilent SureSelect All Exon supplemented with CGD genes - merged: fragment targets ± 20 bp & Padded (Coverage ± 100 bp) & CGD Ensembl ± 20 bp & ASD -21 bp + 20 bp