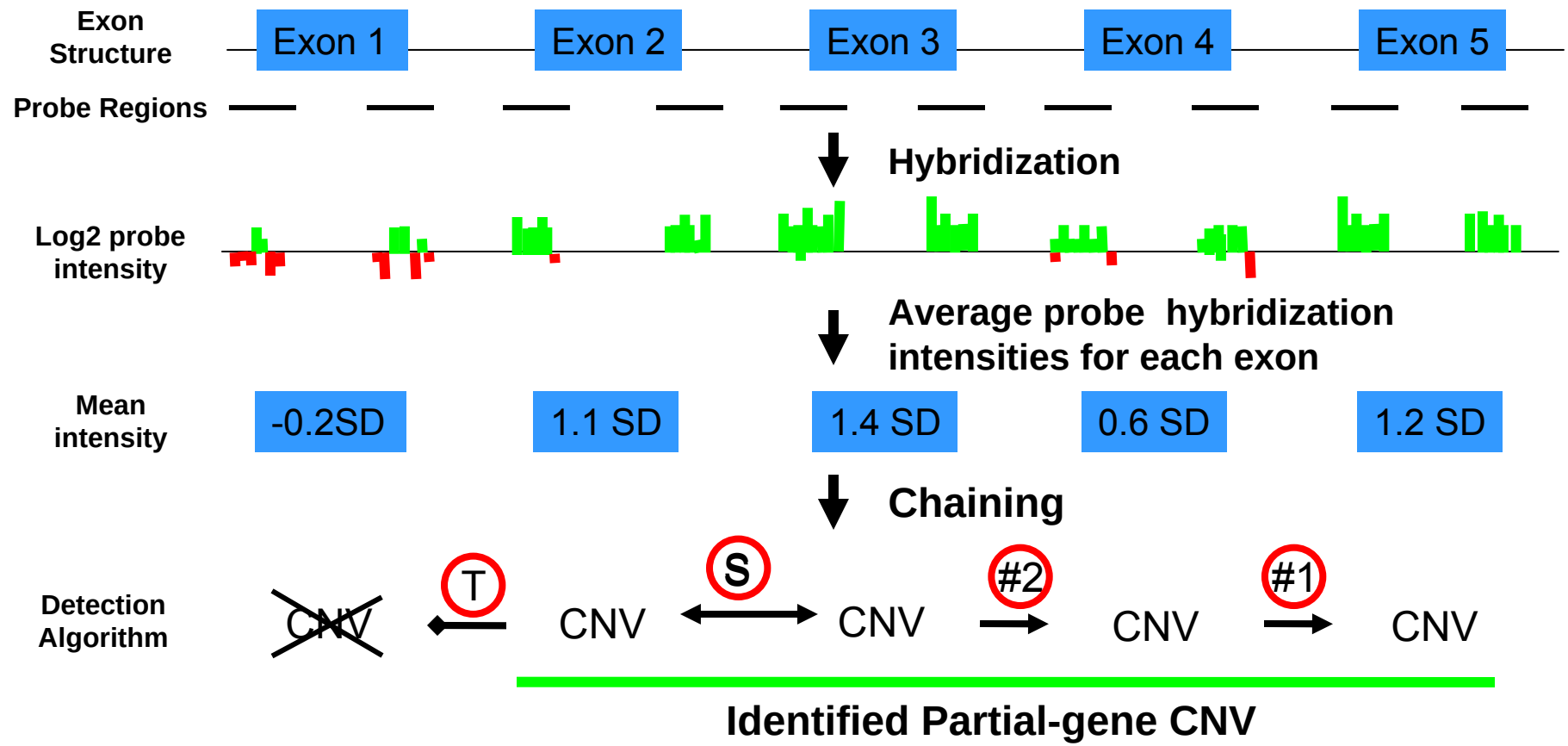
Supplemental Figure 1



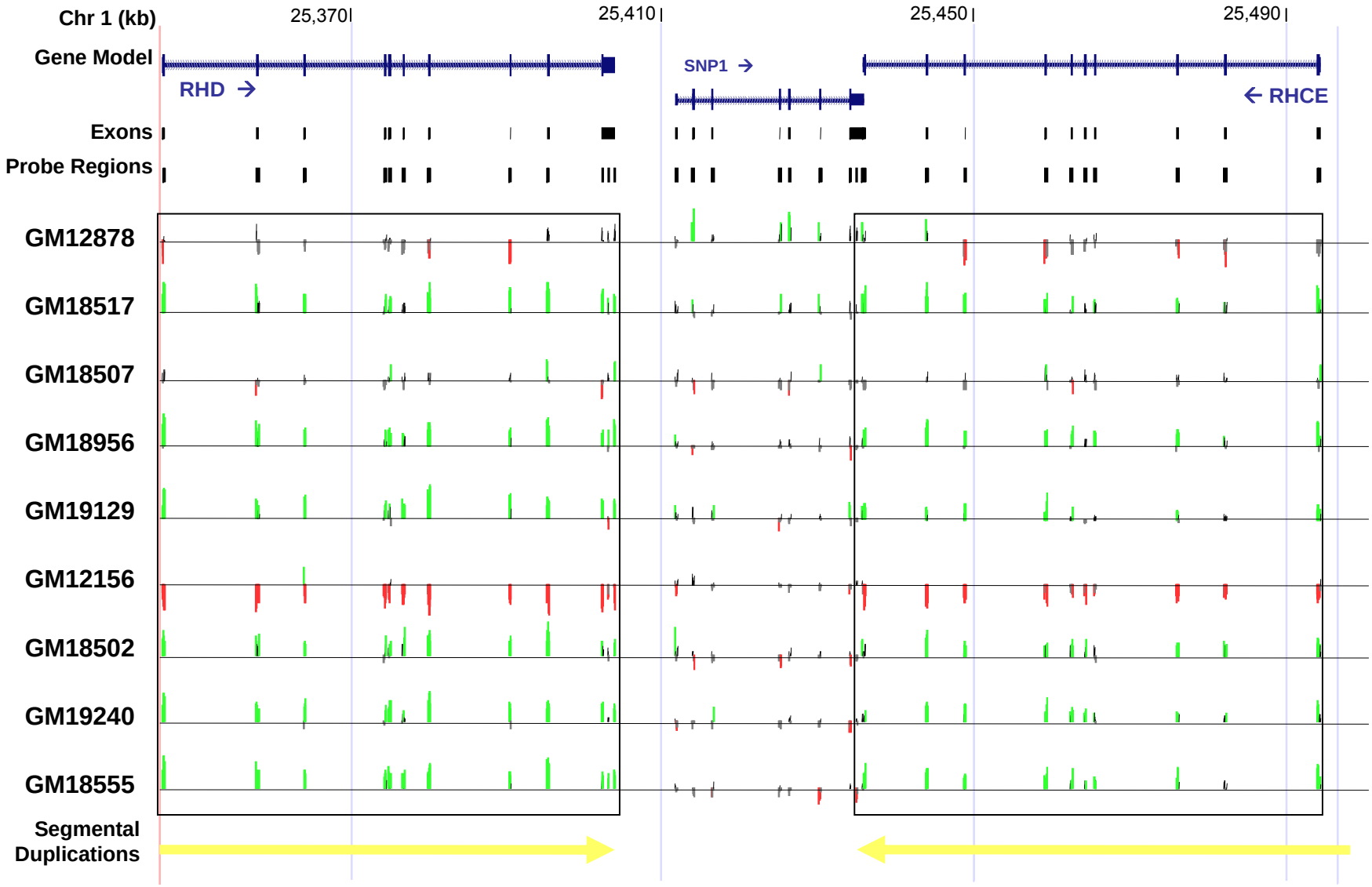
Chaining Rules

- (S)eed: requires ≥1.3 SD with adjacent exon ≥1.0 SD
- Extension Rules: Add adjacent exons to seed if and only if:
  - (#1) ≥1 SD
  - (#2) Up to 2 consecutive exons (0.5-1.0 SD) if flanked by exons ≥1 SD
  - (#3) Or if ≥0.5 if first or last exon and gene >2 exons and all other exons appear CNV
- (T)ermination: if none of the 3 extension criteria are met.

Supplemental Figure 1. CNV detection algorithm

Depicted is a hypothetical example of CNV detection illustrating the rules governing detection and chaining. For each exon, the average hybridization intensity was calculated after trimming the extreme probes (upper and lower deciles). For single exon genes, a threshold of ≥1.3 STD was indicative of CNV. For multi-exon genes, a seeding region was required consisting of two contiguous exons that met the threshold of ≥1.3 STD and ≥1.0 STD, respectively. Further chaining was less stringent to allow for variation in the copy-number responsiveness of an exon's probes. Chain extension continued in three ways: rules #1--3. Rule#1 is simple extension at a high threshold. Rule #2 allowed for dips in relative hybridization intensity (0.5--1.0 SD) as long as the subsequent exons showed strong (≥1 SD) relative hybridization intensity. Rule #3 allowed for the detection of apparent whole-gene CNVs where the terminal exon showed reduced intensity (0.5--1.0 SD). Chaining is terminated if rules #1--3 could not be applied. In this hypothetical example, the seed occurs in exons 2 (≥1.0 SD) and 3 (≥1.3 SD). Continued chain extension occurs to the left by rule #2 for exon 4 and then rule #1 for exon 5. Termination occurs before the inclusion of exon number 1 where the mean relative hybridization intensity is in the opposite direction. Thus, a partial-gene copy number gain of exons 2 through 5 is identified in this example.

Supplemental Figure 2



Supplemental Figure 2. False CNV signatures at RHCE due to paralogous CNV RHD deletion

Depicted is the 110 kb region containing the RHD and RHCE genes along with the observed relative signal intensities of the oligonucleotide probes for all nine individuals. Probes are labeled green and red if the respective relative gain or loss in intensity exceeds a standard deviation for a given individual. The paralogous genes are related by a 50 kb inverted tandem duplication that has 98% nucleotide identity. RHD contains a null mutation due to a deletion of the entire gene that renders 15% of the Caucasian population negative for the protein. The G248 control is heterozygous for this deletion. Within RHCE there are no known common CNVs. Interestingly, RHCE shows a similar albeit weaker pattern of variation relative to RHD in all individuals. These signals on RHCE are due to cross hybridization of RHD DNA onto the highly similar RHCE probes. Gene structures are depicted in blue with RHD and RHCE in transcribed opposite orientations. Probes are labeled red or green if they exceed. Probe regions delineate the regions from which the probes were chosen from vertical bars represent a scale of 5 kb. The SD containing the genes and its inverted orientation is in yellow.