

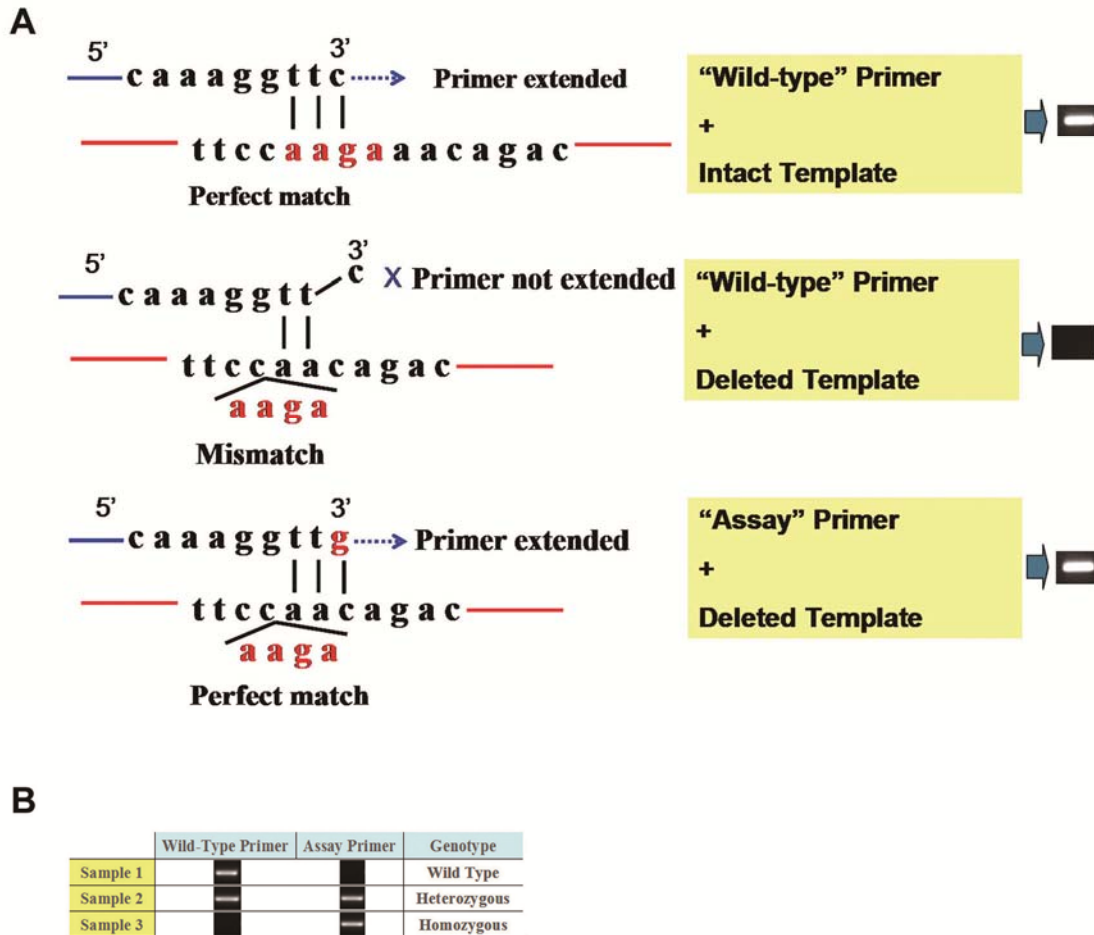
SUPPORTING INFORMATION

|
Scripts and programs used in data analysis:

All computer codes are accessible through the following link:

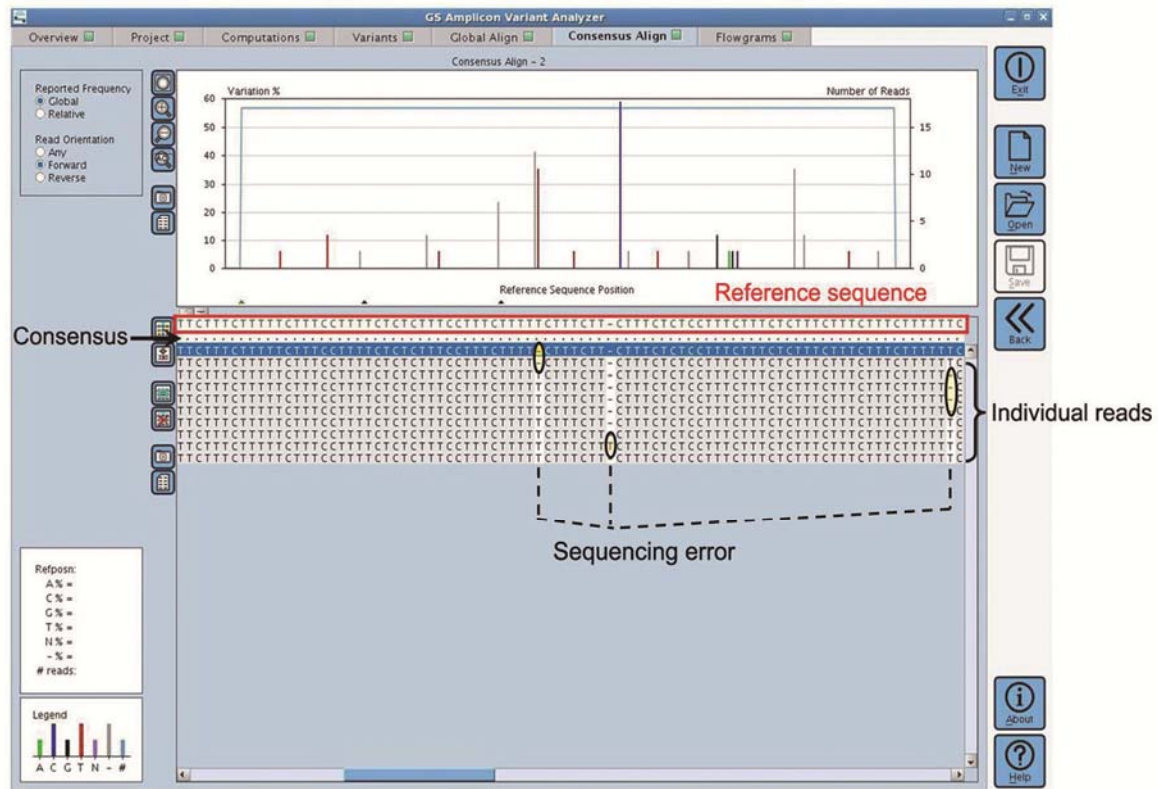
http://chakravarti.igm.jhmi.edu/AravindaChakravartiLab/GS_Junior.html

Supplementary Figure S1



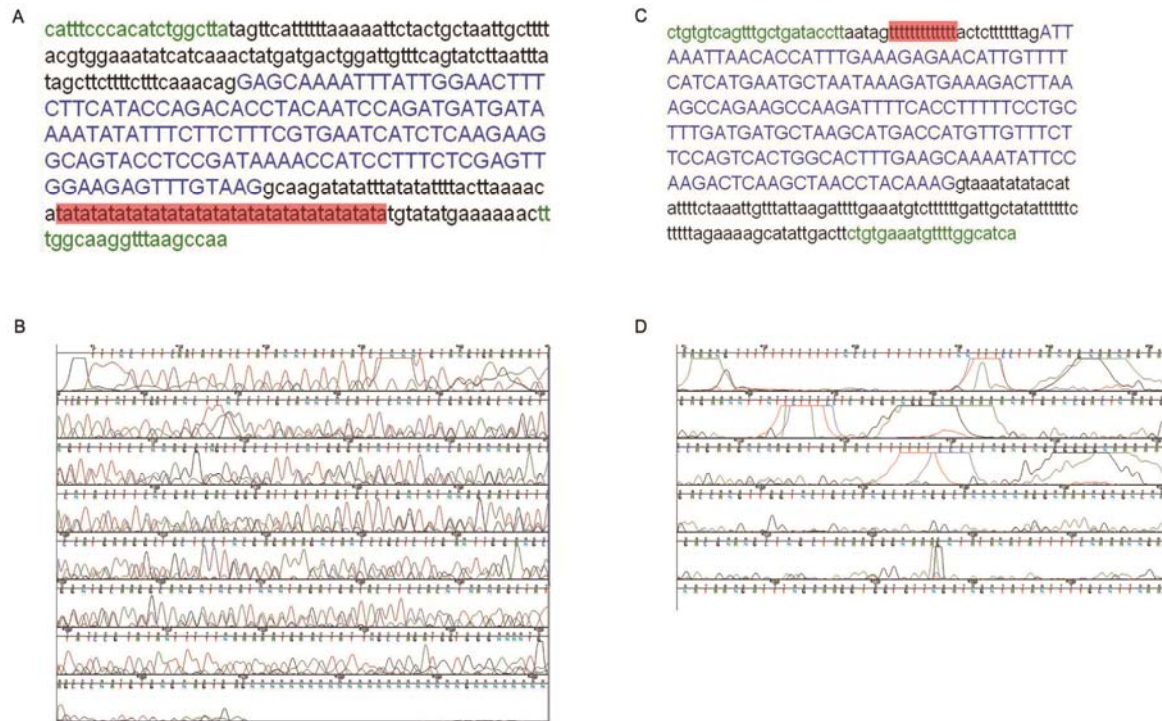
Supp. Figure S1. Allele-specific amplification and detection of PCR products. (A) Schematic representation of the allele-specific PCR method. Wild-type primer forms a perfect match at the 3' end with the intact template on the normal allele but a mismatch with the deleted template on the variant allele. Similarly, the assay primer forms a 3' end match with the deleted allele and 3' end mismatch with the wild type allele. (B) Allele-specific amplification genotypes of each sample after gel electrophoresis.

Supplementary Figure S2



Supp. Figure S2. An example of sequence at a long homopolymeric $(CT)_n$ region. A screenshot of one amplicon sequencing trace after alignment to the reference sequence is shown in the Amplicon Variant Analyzer (AVA). Ten individual reads are shown below with three sequencing errors; the reference sequence is in the first line followed by the consensus sequence.

Supplementary Figure S3



Supp. Figure S3. DNA sequence of two amplicons (A, C) and their corresponding Sanger-sequence chromatograms (B, D). Coding sequences are shown in upper case; PCR amplification and Sanger-sequencing primers are highlighted in green; an AT di-nucleotide repeat and a T homopolymer are shaded in red.

Supplementary Figure S4

SEMA3A

```
Human NP_006071.1 GLANSSSYHTFLLDEERSRLYVGAKDHIFSFDLVNI-KDFQKIVWVPSYT 98  
Chimp_XP_001160811.1 GLANSSSYHTFLLDEERSRLYVGAKDHIFSFDLVNI-KDFQKIVWVPSYT 98  
Mouse NP_033178.2 GLANSSSYHTFLLDEERSRLYVGAKDHIFSFNLVNI-KDFQKIVWVPSYT 98  
Rat NP_059006.1 GLANSSSYHTFLLDEERSRLYVGAKDHIFSFNLVNI-KDFQKIVWVPSYT 98  
Dog_XP_849675.1 GLANSSSYHTFLLDEERSRLYVGAKDHIFSFNLVNI-KDFQKIVWVPSYT 98  
Cow_XP_002686782.1 GLANSSSYHTFLLDEERSRLYVGAKDHIFSFNLVNI-KDFQKIVWVPSYT 98  
Chicken NP_990308.1 GLANSSSYHTFLLDEERSRLYVGAKDHIFSFNLVNI-KEYQKIVWVPSHS 98  
Zebrafish NP_571136.1 GLANSSAYHTFLLDEERGRLLVFGAKDHLVLSFNLDVINDMQQLISWPSSPS 100  
*****:*****:*:*****:***:*:* : * * * * * :
```

↳ S54G

```
Human NP_006071.1 VITFARSHPAMYNPVFPMNRPVIVIKTDVNYQFTQIVVDRVDAEDGGQYDV 448  
Chimp_XP_001160811.1 VITFARSHPAMYNPVFPMNRPVIVIKTDVNYQFTQIVVDRVDAEDGGQYDV 448  
Mouse NP_033178.2 VITFARSHPAMYNPVFPIINRPIIMIKTDVNYQFTQIVVDRVDAEDGGQYDV 448  
Rat NP_059006.1 VITFARSHPAMYNPVFPIINRPIIMIKTDVNYQFTQIVVDRVDAEDGGQYDV 448  
Dog_XP_849675.1 VITFARSHPAMYNPVFPIINRPIIMIKTDVNYQFTQIVVDRVDAEDGGQYDV 448  
Cow_XP_002686782.1 VITFARSHPAMYNPVFPIINRPIIMIKTDVNYQFTQIVVDRVDAEDGGQYDV 448  
Chicken NP_990308.1 VITFARSHPAMYNPVFPIINRPIIMIKTDVNYQFTQIVVDRVDAEDGGQYDV 448  
Zebrafish NP_571136.1 VITFARSHPAMYNPVFPIINHPPIIKTDVDYQFTQIVVDRVDAEDGGQYDV 449  
*****:*****:*:*****:*****:*****:
```

↳ V435I

SEMA3C

```
Human NP_006370.1 EDGPETHFDELEDVFLLETDPNRTTLVYGIFFTSSSVFKGSAVCVYHLSD 344  
Chimp_XP_527801.2 EDGPETHFDELEDVFLLETDPNRTTLVYGIFFTSSSVFKGSAVCVYHLSD 344  
Mouse NP_038685.3 EDGPETHFDELEDVFLLETDPNRTTLVYGIFFTSSSVFKGSAVCVYHLSD 344  
Rat NP_001100048.1 EDGPETHFDELEDVFLLETDPNRTTLVYGIFFTSSSVFKGSAVCVYHLSD 344  
Dog_XP_533139.2 EDGPETHFDELEDVFLLETDPNRTTLVYGIFFTSSSVFKGSAVCVYHLSD 344  
Cow NP_001094552.1 EDGPETHFDELEDVFLLEMDNPRRTLAVYGIFFTSSSVFKGSAVCVYHFS 344  
Chicken NP_989574.1 EDGTETYFDELEDVFLLETDPNRTTLVYGIFFTSSSVFKGSAVCVYHLSD 344  
Zebrafish_XP_692847.2 EDGTETHFDELENVFLLETDPHFKGLLIFGVFTSTSSVFRGSAVCVYHMD 400  
***:*:*****:***** *:*: *:*:*:*:*:*:*:*:***:***:***:
```

↳ V337M

SEMA3D

```
Human NP_689967.2 LTYKDLLLSNSCIPFLGSEGLDFQTLTLLDEERGRLLLGAKDHIFLLSLV 96  
Chimp_XP_527805.2 LTYKDLLLSNSCIPFLGSEGLDFQTLTLLDEERGRLLLGAKDHIFLLSLV 96  
Mouse NP_083158.3 LTYKDLLLSNTCIPFLGSEGLDFQTLTLLDEERGIILLGAKDHVFLSLV 96  
Rat NP_001098103.1 LTYKDLLLSNTCIPFLGSEGLDFQTLTLLDEERGIILLGAKDHVFLNLV 96  
Dog_XP_533138.2 LTYKDLLLSNSCIPFLGSEGLDFQTLTLLDEERGRLLVVGAKDHIFLLSLV 96  
Cow_XP_616548.3 LTYKDLLLSNSCIPFLGSEGLDFQTLTLLDEERGRLLLGAKDHIFLLSLV 96  
Chicken NP_990704.1 LSYKDLLLSNSCIPFLGSEGLDFRTLLTLLDEERGRLLVVGAKDHIFLLNLV 84  
Zebrafish NP_571123.1 LGYKDLIHSRVSVPFTGSEGHQFTVLLDEERSRLLLGAKDHVYLLDPD 100  
* * * * * : * . : ** * * . * * . : * : * * * * * . * : * * * * * : * *
```

↳ S65P

```
Human NP_689967.2 DEKVIFGIEFNSTFLECI PKSQQATIKWYIQRSGDEHREELKPDERI IKT 646  
Chimp_XP_527805.2 DEKVIFGIEFNSTFLECI PKSQQATIKWYIQRSGDEHREELKPDERI IKT 646  
Mouse NP_083158.3 DEKVIFGIEFNSTFLECI PKSQQASVEWYIQRSGDEHREELKPDERI IKT 646  
Rat NP_001098103.1 DEKVIFGIEFNSTFLECI PKSQQASVEWYIQRSGDEHREELKPDERI IKT 646  
Dog_XP_533138.2 DEKVIFGIEFNSTFLECI PKSQQASIKWYIQRSGDEHREELKPDERI IKT 646  
Cow_XP_616548.3 DEKVIFGIEFNSTFLECI PKSQQASVIRWYIQRSGDEHREELKPDERI IKT 645  
Chicken NP_990704.1 DEKVIFGIEFNSTFLECI PKSQQASIRWYIQRSGEHEHREELKADERI IKT 634  
Zebrafish NP_571123.1 EEKVLVGVESNSSFLECVSKSQQALIRWFLKPGVDHRQEIKPDERVIT 649  
:***:*:* * * : * * * * * : * * * * * : * * * * * : * *
```

↳ P615T

```
Human NP_689967.2 HEEGVVKDLLAESRLRYKDYIQLSSPNFSLDQYCEQMWHEKRRQRNKG 746  
Chimp_XP_527805.2 HEEGVVKDLLAESRLRYKDYIQLSSPNFSLDQYCEQMWHEKRRQRNKG 746  
Mouse NP_083158.3 YQEGQVKDLLAESRLRYKDYIQLSSPNFSLDQYCEQMWWYKRRQRNKG 746  
Rat NP_001098103.1 YQEGQVKDLLAESRLRYKDYIQLSSPNFSLDQYCEQMWWYKRRQRNKG 746  
Dog_XP_533138.2 HEEGVVKDLLAESRLRYKDYIQLSSPNFSLDQYCEQMWHEKRRQRNKG 746  
Cow_XP_616548.3 HEEGVVKDLLAESRLRYKDYIQLSSPNFSLDQYCEQMWHEKRRQRNKG 745  
Chicken NP_990704.1 DEEGRVLDLLESRLRYKDYIQLVSSPSFSLDEYCEQMWHEKRRQRNKG 734  
Zebrafish NP_571123.1 R-ESSENPAVTEPRQRYKDYLRMLSGPARSLDEYCEQMWHEK- QKQKG 747  
* . . : * . * * * * * : * * * * * * * * * * * : * * * * *
```

↳ K701Q

Supp. Figure S4. Protein sequence alignment of SEMA3A, SEMA3C and SEMA3D in human, chimp, mouse, rat, dog, cow, chicken and zebrafish. The position of each missense variant identified in HSCR cases is shown; asterisks denote completely conserved sites while dots indicate conservative substitutions.