# UpSet: Visualization of Intersecting Sets

## Supplementary Material

Alexander Lex
alex@seas.harvard.edu
Harvard University

Nils Gehlenborg
nils@hms.harvard.edu
Harvard Medical School

Hendrik Strobelt
strobelt@seas.harvard.edu
Harvard University

Romain Vuillemot
romain_vuillemot@hks.harvard.edu
Harvard University

Hanspeter Pfister
pfister@seas.harvard.edu
Harvard University

IEEE Transactions on Visualization and Computer Graphics—InfoVis 2014

# Supplementary Figure 1

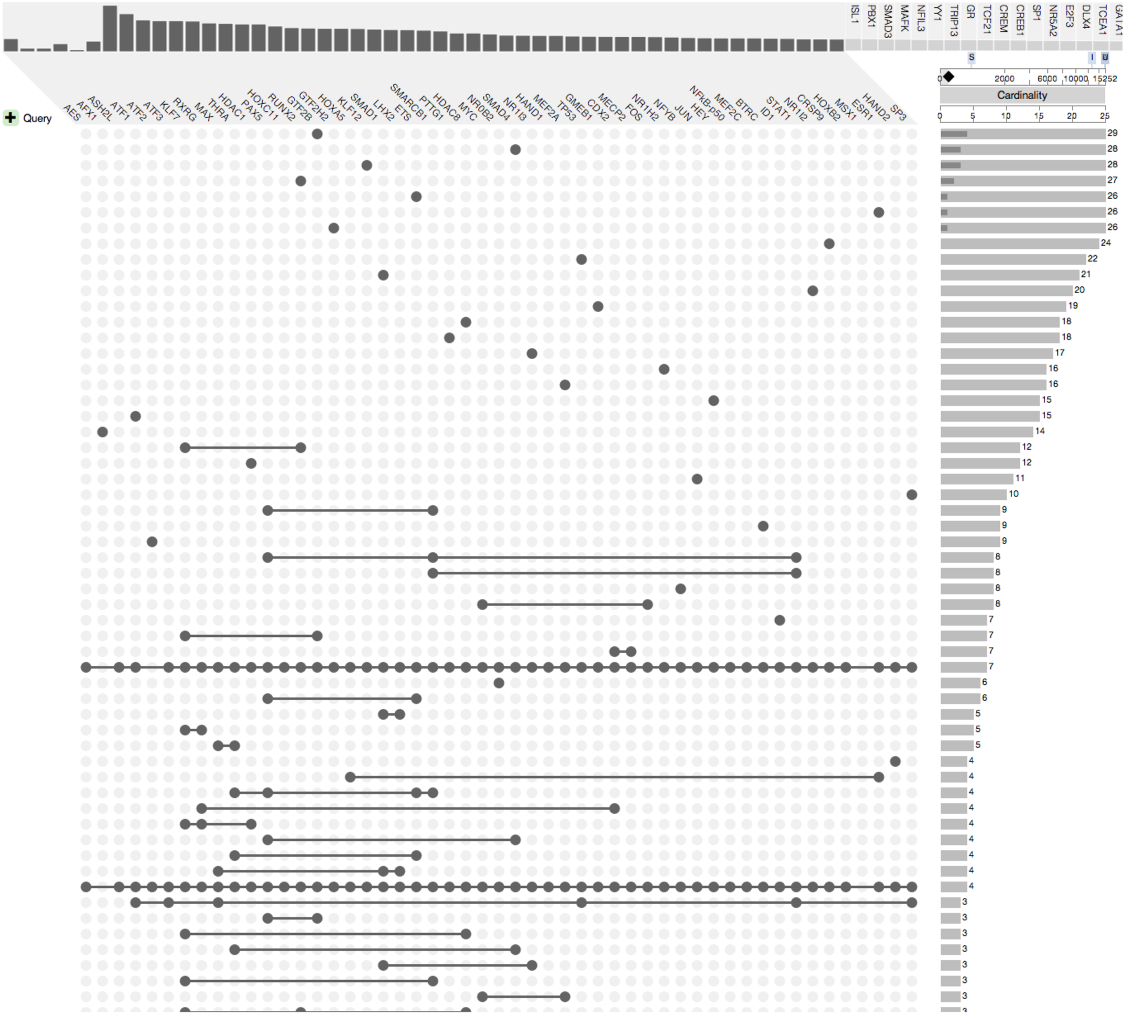


**Figure S1:** Scalability demonstration showing a dataset of protein-compound interactions. The proteins are the sets, the compounds the elements. We show co-interactions for 51 proteins. It is evident that some compounds are not specific and interact with almost all proteins.

# Supplementary Figure 2

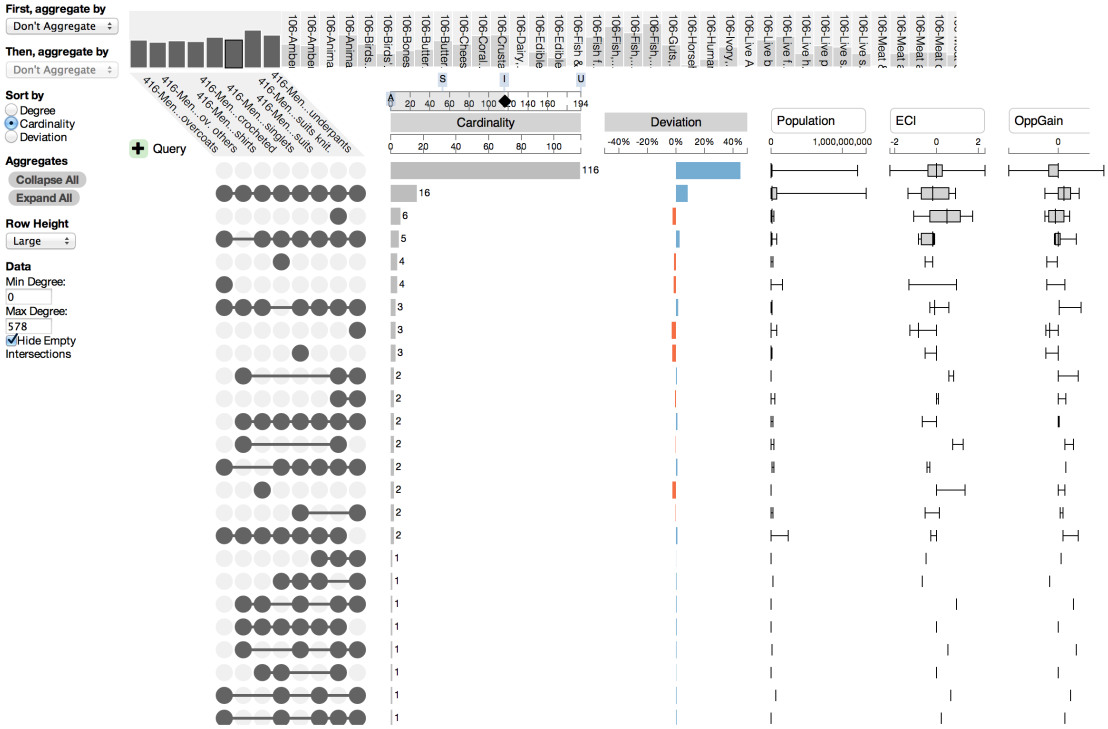
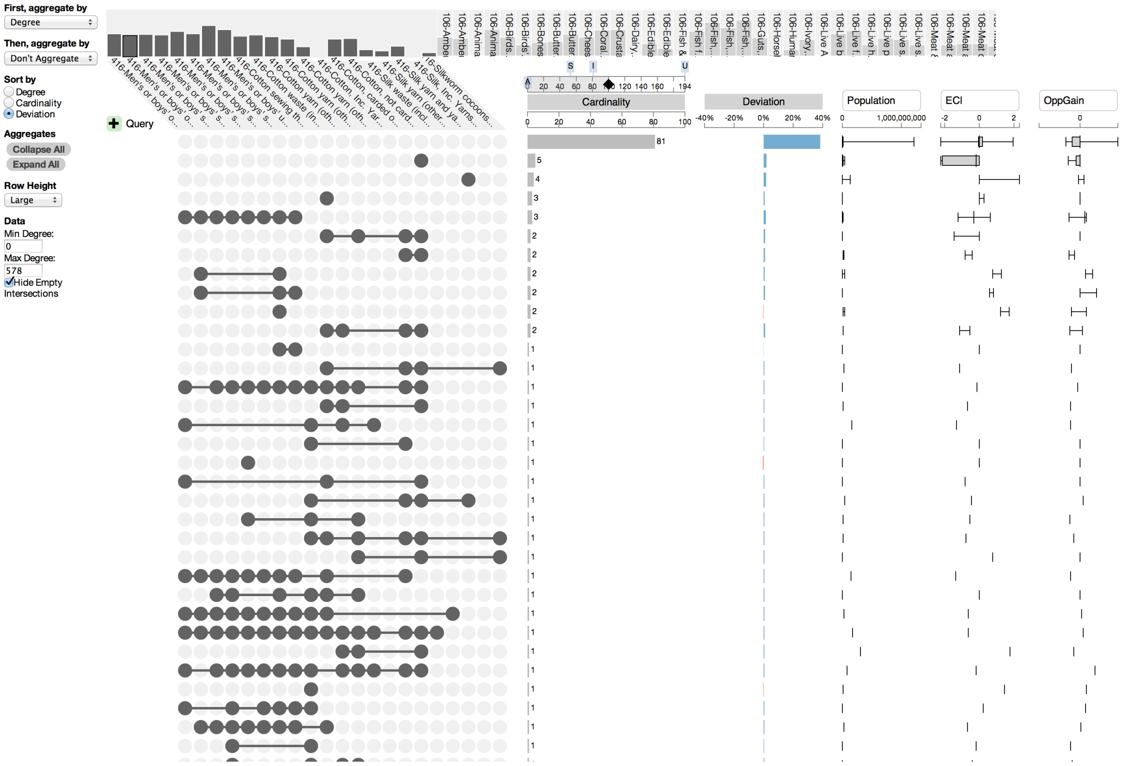

First, aggregate by: Don't Aggregate
Then, aggregate by: Don't Aggregate
Sort by: ○ Degree ● Cardinality ○ Deviation
Row Height: Large
Data
Min Degree: 0
Max Degree: 578
☑ Hide Empty Intersections

**Figure S2:** Selection of 8 products, with intersections ordered by cardinality. The largest intersection of products are countries that do not export any of the selected products (first row). The second largest one is countries that are exporting all products together, suggesting that countries that produce a specific type of men's textiles often also produce similar other products.

# Supplementary Figure 3


First, aggregate by: Degree

Then, aggregate by: Don't Aggregate

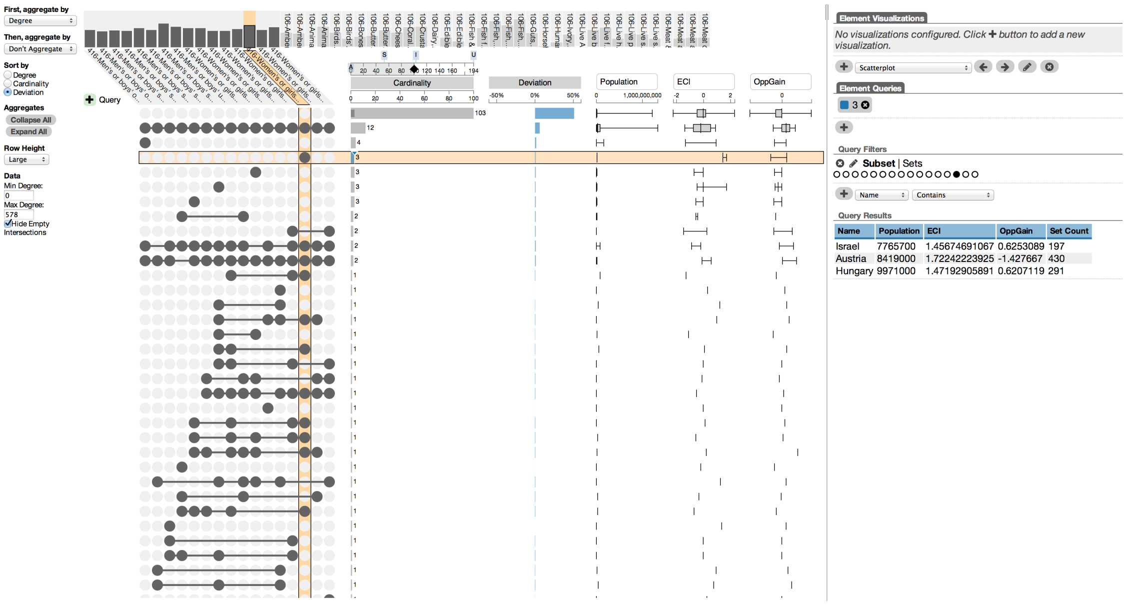
Sort by
○ Degree
○ Cardinality
● Deviation

Row Height: Large

Data
Min Degree: 0
Max Degree: 578
☑ Hide Empty Intersections

**Figure S3:** Men's textile products combined with other products related to *Cotton* (8 products) and *Silk* (5 products). The intersections are sorted by cardinality, revealing that there is no overlap between these products, leading to the conclusion that there is no vertical integration between these products, i.e., countries that produce cotton and silk raw materials do not typically manufacture men's textile products.

# Supplementary Figure 4



First, aggregate by: Degree

Then, aggregate by: Don't Aggregate

Sort by
○ Degree
○ Cardinality
● Deviation

Aggregates: Collapse All, Expand All

Row Height: Large

Data
Min Degree: 0
Max Degree: 578
☑ Hide Empty Intersections

No visualizations configured. Click ✚ button to add a new visualization.

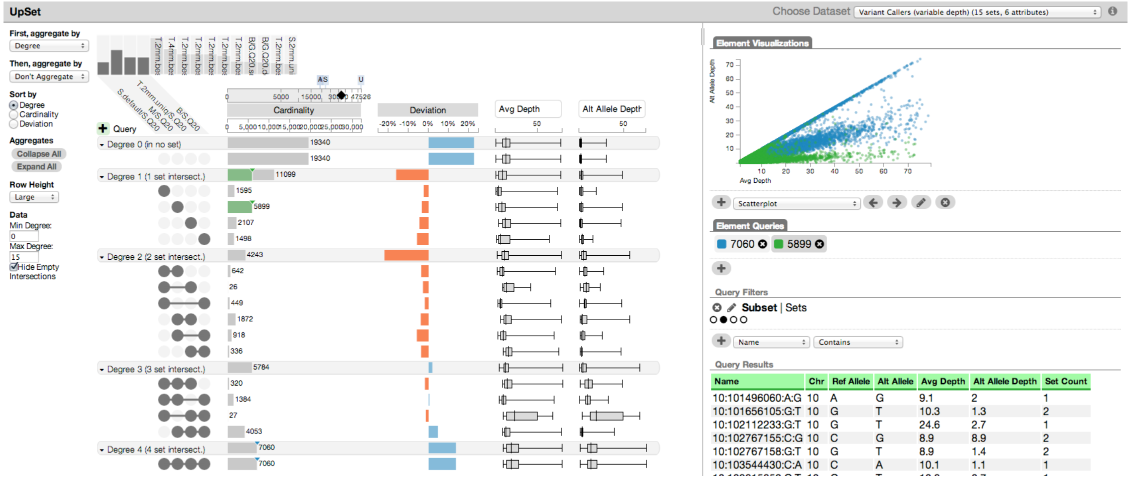
Scatterplot

Name: Contains

| Name | Population | ECI | OppGain | Set Count |
|---|---|---|---|---|
| Israel | 7765700 | 1.45674691067 | 0.6253089 | 197 |
| Austria | 8419000 | 1.72242223925 | -1.427667 | 430 |
| Hungary | 9971000 | 1.47192905891 | 0.6207119 | 291 |

**Figure S4:** Countries exporting both men's and women's textile products. Notice the outlier highlighted in orange. The three countries (Israel, Austria and Hungary) in this intersection produce only a specific type of women's clothing and have a notably higher Economic Complexity Index (ECI) than other countries that manufacture textile products. This suggests that some products this product category ("women's or girls' slips, petticoats, briefs, panties, night-dresses, pajamas, negligees, bathrobes, dressing gowns and similar articles, knitted or crocheted") require techniques available mainly in developed and diversified countries.

# Supplementary Figure 5



First, aggregate by: Degree

Then, aggregate by: Don't Aggregate

Sort by: ● Degree ○ Cardinality ○ Deviation

Row Height: Large

Data

Min Degree: 0

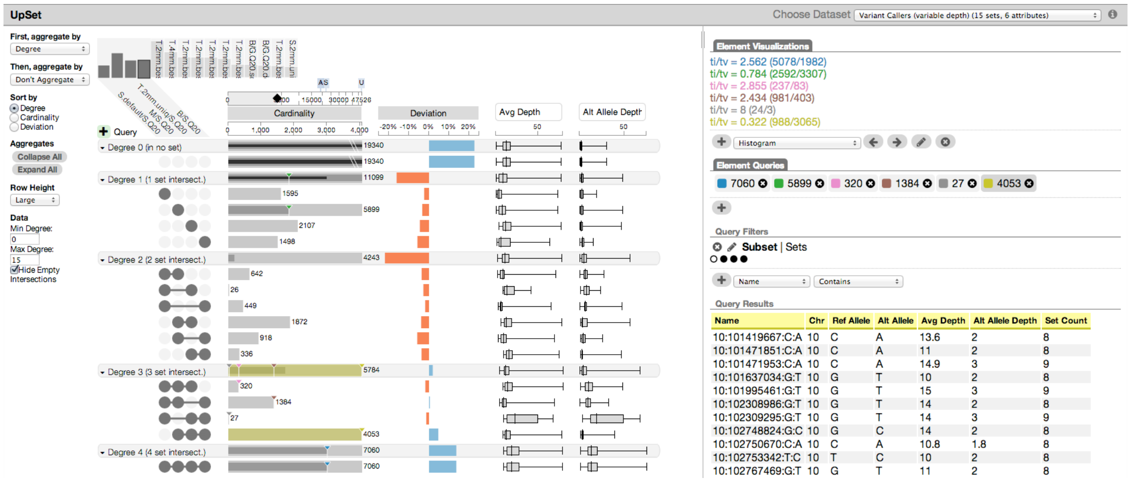
Max Degree: 15

☑ Hide Empty Intersections

| Name | Chr | Ref Allele | Alt Allele | Avg Depth | Alt Allele Depth | Set Count |
|---|---|---|---|---|---|---|
| 10:101496060:A:G | 10 | A | G | 9.1 | 2 | 1 |
| 10:101656105:G:T | 10 | G | T | 10.3 | 1.3 | 2 |
| 10:102112233:G:T | 10 | G | T | 24.6 | 2.7 | 1 |
| 10:102767155:C:G | 10 | C | G | 8.9 | 8.9 | 2 |
| 10:102767158:G:T | 10 | G | T | 8.9 | 1.4 | 2 |
| 10:103544430:C:A | 10 | C | A | 10.1 | 1.1 | 1 |

**Figure S5:** Scatterplot showing the correlation of *Average Depth* and *Alternative Allele Depth* for two queries. The point cloud along the diagonal corresponds to variants that were found in all sequence fragments at the given position and most likely are present in both copies of the genome in the cells of the human individual (homozygous variants). The second major point cloud, between the *x* axis and the diagonal, corresponds to variants that were found in around half of all sequence fragments at the given position. These most likely are present in one of the two copies of the genome in the cells of the human individual (heterozygous variants). Finally, the last major point cloud (corresponding to the green query), consists of variants found in much fewer than half of all sequence fragments at the given position, which is an indicator that they might not be real. These are the calls made predominantly by the presumably inferior caller M/S.Q20, highlighted in green in the set view.

# Supplementary Figure 6



**UpSet**

Choose Dataset: Variant Callers (variable depth) (15 sets, 6 attributes)

**First, aggregate by:** Degree

**Then, aggregate by:** Don't Aggregate

**Sort by**
● Degree
○ Cardinality
○ Deviation

**Row Height:** Large

**Data**
Min Degree: 0
Max Degree: 15
☑ Hide Empty Intersections

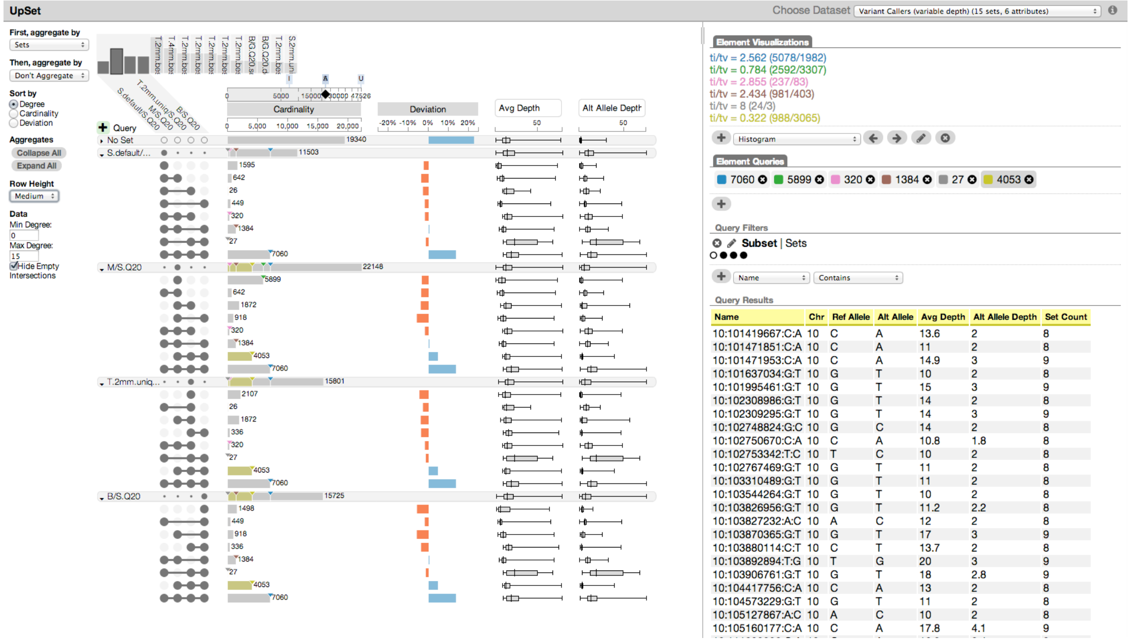
**Element Visualizations**

ti/tv = 2.562 (5078/1982)
ti/tv = 0.784 (2592/3307)
ti/tv = 2.855 (237/83)
ti/tv = 2.434 (981/403)
ti/tv = 8 (24/3)
ti/tv = 0.322 (988/3065)

| Name | Chr | Ref Allele | Alt Allele | Avg Depth | Alt Allele Depth | Set Count |
|---|---|---|---|---|---|---|
| 10:101419667:C:A | 10 | C | A | 13.6 | 2 | 8 |
| 10:101471851:C:A | 10 | C | A | 11 | 2 | 8 |
| 10:101471953:C:A | 10 | C | A | 14.9 | 3 | 9 |
| 10:101637034:G:T | 10 | G | T | 10 | 2 | 8 |
| 10:101995461:G:T | 10 | G | T | 15 | 3 | 9 |
| 10:102308986:G:T | 10 | G | T | 14 | 2 | 8 |
| 10:102309295:G:T | 10 | G | T | 14 | 3 | 9 |
| 10:102748824:G:C | 10 | G | C | 14 | 2 | 8 |
| 10:102750670:C:A | 10 | C | A | 10.8 | 1.8 | 8 |
| 10:102753342:T:C | 10 | T | C | 10 | 2 | 8 |
| 10:102767469:G:T | 10 | G | T | 11 | 2 | 8 |

**Figure S6:** Intersections aggregated by degree and variant frequency plot showing transition/transversion ratios for the previous queries, as well as all 3-set intersections. The large size of the intersection corresponding to the yellow query stands out in the set view.

# Supplementary Figure 7



Choose Dataset: Variant Callers (variable depth) (15 sets, 6 attributes)

First, aggregate by: Sets

Then, aggregate by: Don't Aggregate

Sort by
● Degree
○ Cardinality
○ Deviation

Row Height: Medium

Data
Min Degree: 0
Max Degree: 15
☑ Hide Empty Intersections

ti/tv = 2.562 (5078/1982)
ti/tv = 0.784 (2592/3307)
ti/tv = 2.855 (237/83)
ti/tv = 2.434 (981/403)
ti/tv = 8 (24/3)
ti/tv = 0.322 (988/3065)

Element Queries: 7060, 5899, 320, 1384, 27, 4053

Query Filters
Subset | Sets
○●●●

Name / Contains

| Name | Chr | Ref Allele | Alt Allele | Avg Depth | Alt Allele Depth | Set Count |
|---|---|---|---|---|---|---|
| 10:101419667:C:A | 10 | C | A | 13.6 | 2 | 8 |
| 10:101471851:C:A | 10 | C | A | 11 | 2 | 8 |
| 10:101471953:C:A | 10 | C | A | 14.9 | 3 | 9 |
| 10:101637034:G:T | 10 | G | T | 10 | 2 | 8 |
| 10:101995461:G:T | 10 | G | T | 15 | 3 | 9 |
| 10:102308986:G:T | 10 | G | T | 14 | 2 | 8 |
| 10:102309295:G:T | 10 | G | T | 14 | 3 | 9 |
| 10:102748824:G:C | 10 | G | C | 14 | 2 | 8 |
| 10:102750670:C:A | 10 | C | A | 10.8 | 1.8 | 8 |
| 10:102753342:T:C | 10 | T | C | 10 | 2 | 8 |
| 10:102767469:G:T | 10 | G | T | 11 | 2 | 8 |
| 10:103310489:G:T | 10 | G | T | 11 | 2 | 8 |
| 10:103544264:G:T | 10 | G | T | 10 | 2 | 8 |
| 10:103826956:G:T | 10 | G | T | 11.2 | 2.2 | 8 |
| 10:103827232:A:C | 10 | A | C | 12 | 2 | 8 |
| 10:103870365:G:T | 10 | G | T | 17 | 3 | 9 |
| 10:103880114:C:T | 10 | C | T | 13.7 | 2 | 8 |
| 10:103892894:T:G | 10 | T | G | 20 | 3 | 9 |
| 10:103906761:G:T | 10 | G | T | 18 | 2.8 | 9 |
| 10:104417756:C:A | 10 | C | A | 13 | 2 | 8 |
| 10:104573229:G:T | 10 | G | T | 11 | 2 | 8 |
| 10:105127867:A:C | 10 | A | C | 10 | 2 | 8 |
| 10:105160177:C:A | 10 | C | A | 17.8 | 4.1 | 9 |

**Figure S7:** Intersections aggregated by set and variant frequency plot showing transition/transversion ratios for the queries from Supplementary Figure S2. This figure illustrates that almost all of the the variants called by the combination S.default/S.Q20 were also found by the other combinations selected here. All variants called by S.default/S.Q20 and at least two of the other combinations are either of high quality with transition/transversion ratios close to 2.7 (pink, brown). In one case (gray) the number of variants called is too small to yield a reliable estimate.