# Interactive software for the integrated analysis and identification of rare and undiagnosed diseases using NGS data

**Stephan Pabinger**, Denis Katic, Ana Krolo, Tatjana T. Hirschmugl, Kaan Boztug, Albert Kriegner, Klemens Vierlinger

Austrian Institute of Technology AIT

Platomics

CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences

stephan.pabinger@ait.ac.at | @tadkeys

# NGS for rare / undiagnosed diseases



HiSeq 3000/HiSeq 4000 Systems

NextSeq

MiSeq
PGM
Proton

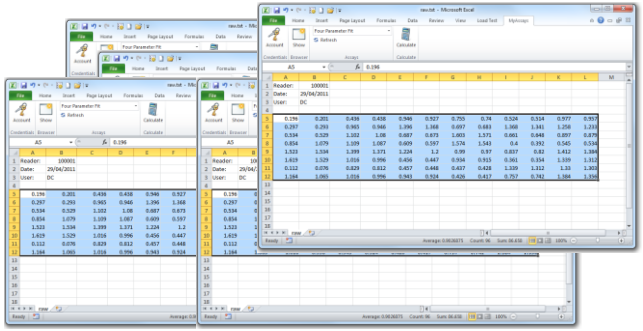
# Output

| VarChr | VarStart | VarEnd | DNAChange | VarType | VarClass | VarPercenta | RefCov | VarCov | dbSnpId | Transcript | VarId | TotalCov |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| chr13 | 32890572 | 32890572 | G>A | SNP | R | 100 | 0 | 184 | rs1799943 | NM_000059 | 16 | 184 |
| chr13 | 32890572 | 32890572 | G>A | SNP | R | 100 | 0 | 249 | rs1799943 | NM_000059 | 37 | 249 |
| chr13 | 32890572 | 32890572 | G>A | SNP | R | 99,62 | 0 | 530 | rs1799943 | NM_000059 | 37 | 532 |
| chr13 | 32890572 | 32890572 | G>A | SNP | R | 100 | 0 | 98 | rs1799943 | NM_000059 | 13 | 98 |
| chr13 | 32890572 | 32890572 | G>A | SNP | R | 99,62 | 0 | 1294 | rs1799943 | NM_000059 | 40 | 1299 |
| chr13 | 32899388 | 32899388 | A>C | SNP | R | 99,37 | 0 | 631 | rs11571610 | NM_000059 | 23 | 635 |
| chr13 | 32900933 | 32900933 | T>A | SNP | R | 99,76 | 0 | 1681 | rs3752451 | NM_000059 | 14 | 1685 |
| chr13 | 32900933 | 32900933 | T>A | SNP | R | 99,13 | 0 | 227 | rs3752451 | NM_000059 | 13 | 229 |
| chr13 | 32900933 | 32900933 | T>A | SNP | R | 99,81 | 0 | 1584 | rs3752451 | NM_000059 | 15 | 1587 |
| chr13 | 32900933 | 32900933 | T>A | SNP | R | 99,44 | 0 | 536 | rs3752451 | NM_000059 | 13 | 539 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,7 | 0 | 673 | rs206073 | NM_000059 | 7 | 675 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,89 | 0 | 936 | rs206073 | NM_000059 | 10 | 937 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,44 | 0 | 530 | rs206073 | NM_000059 | 18 | 533 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 650 | rs206073 | NM_000059 | 18 | 650 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 543 | rs206073 | NM_000059 | 16 | 543 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,74 | 0 | 780 | rs206073 | NM_000059 | 10 | 782 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 463 | rs206073 | NM_000059 | 16 | 463 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,81 | 0 | 530 | rs206073 | NM_000059 | 9 | 531 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 636 | rs206073 | NM_000059 | 14 | 636 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,85 | 0 | 645 | rs206073 | NM_000059 | 19 | 646 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 109 | rs206073 | NM_000059 | 9 | 109 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 72 | rs206073 | NM_000059 | 14 | 72 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 107 | rs206073 | NM_000059 | 7 | 107 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,89 | 0 | 920 | rs206073 | NM_000059 | 24 | 921 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 783 | rs206073 | NM_000059 | 16 | 783 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,88 | 0 | 868 | rs206073 | NM_000059 | 9 | 869 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,76 | 0 | 842 | rs206073 | NM_000059 | 16 | 844 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,87 | 0 | 777 | rs206073 | NM_000059 | 17 | 778 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 671 | rs206073 | NM_000059 | 16 | 671 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 1272 | rs206073 | NM_000059 | 8 | 1272 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 1341 | rs206073 | NM_000059 | 15 | 1341 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 791 | rs206073 | NM_000059 | 8 | 791 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,84 | 0 | 613 | rs206073 | NM_000059 | 20 | 614 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,72 | 0 | 351 | rs206073 | NM_000059 | 16 | 352 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 99,72 | 0 | 702 | rs206073 | NM_000059 | 9 | 704 |
| chr13 | 32905265 | 32905265 | G>A | SNP | R | 100 | 0 | 828 | rs206073 | NM_000059 | 8 | 828 |

# Finding the “needle in the haystack“

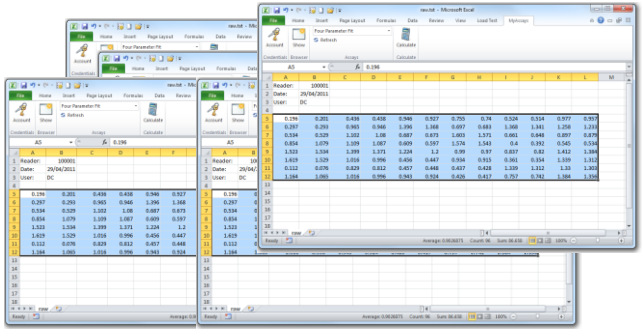
# Software design

Instead of lots of Excel files



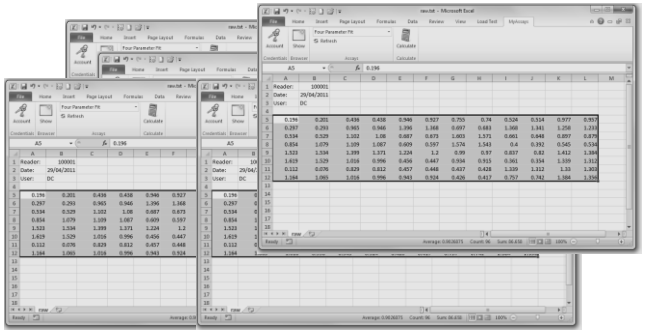
# Software design

Instead of lots of Excel files → Results stored in one place

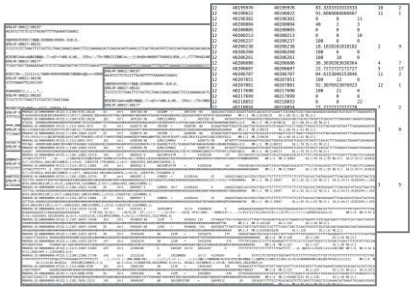


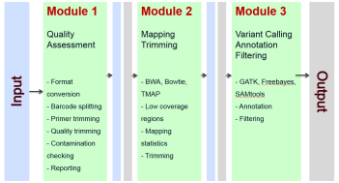
# Software design

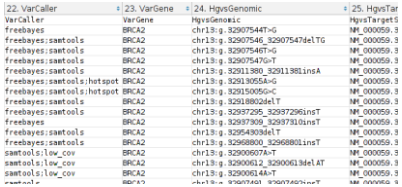
Instead of lots of Excel files → Results stored in one place



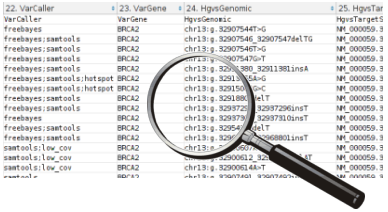
Raw data → Analysis → List of variants → Identification





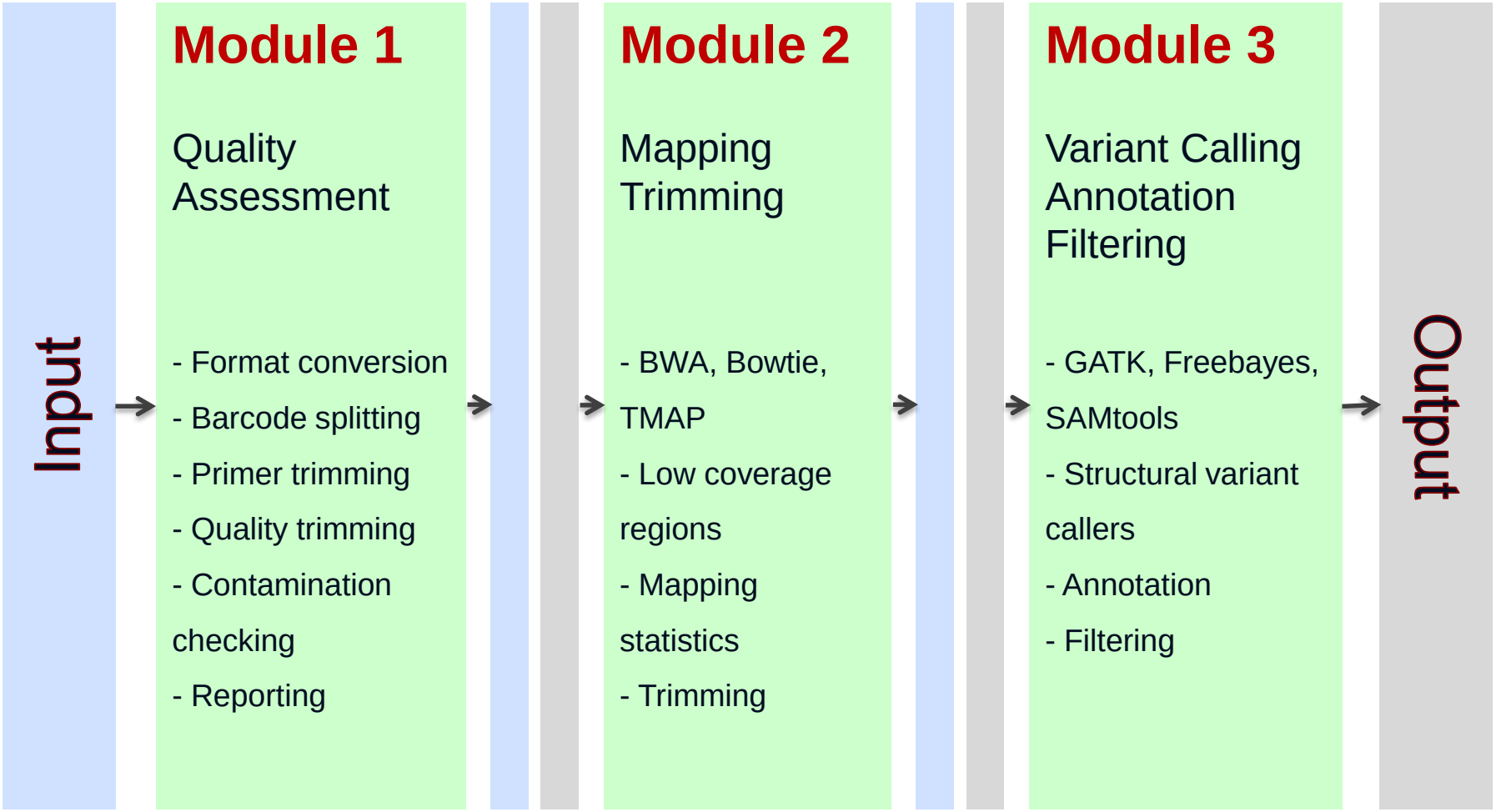



| 22. VarCaller | 23. VarGene | 24. HgvsGenomic | 25. HgvsTar |
|---|---|---|---|
| VarCaller | VarGene | HgvsGenomic | HgvsTargetS |
| freebayes | BRCA2 | chr13:g.32907544T>G | NM_000059.3 |
| freebayes;samtools | BRCA2 | chr13:g.32907546_32907547delTG | NM_000059.3 |
| freebayes;samtools | BRCA2 | chr13:g.32907546T>G | NM_000059.3 |
| freebayes;samtools | BRCA2 | chr13:g.32907547G>T | NM_000059.3 |
| freebayes;samtools | BRCA2 | chr13:g.32911380_32911381insA | NM_000059.3 |
| freebayes;samtools;hotspot | BRCA2 | chr13:g.32913055A>G | NM_000059.3 |
| freebayes;samtools;hotspot | BRCA2 | chr13:g.32915005G>C | NM_000059.3 |
| freebayes;samtools | BRCA2 | chr13:g.32918802delT | NM_000059.3 |
| freebayes;samtools | BRCA2 | chr13:g.32937295_32937296insT | NM_000059.3 |
| freebayes | BRCA2 | chr13:g.32937309_32937310insT | NM_000059.3 |
| freebayes;samtools | BRCA2 | chr13:g.32954303delT | NM_000059.3 |
| freebayes;samtools | BRCA2 | chr13:g.32968800_32968801insT | NM_000059.3 |
| samtools;low_cov | BRCA2 | chr13:g.32900607A>T | NM_000059.3 |
| samtools;low_cov | BRCA2 | chr13:g.32900612_32900613delAT | NM_000059.3 |
| samtools;low_cov | BRCA2 | chr13:g.32900614A>T | NM_000059.3 |



# Multistep Application

Accepts input from all big NGS technologies (Illumina, Ion Torrent, 454 ...)



# Reproducibility & Configuration

## Logging

- Complete log of all used tools, references, annotation databases, and versions

## Storing

- Storage of output and input data → Run and re-run analyses

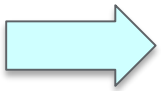
## Accessing

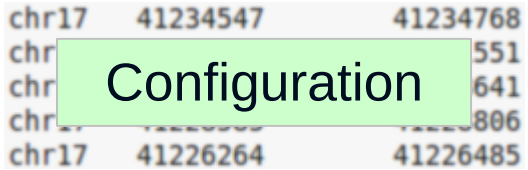
- Get all data from all samples at any time

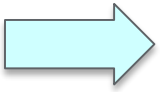
## Configuring

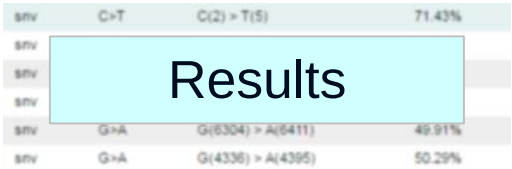
- Specify exactly which genes/regions should be analyzed

BMPR2
ELN
SCNN1A







# Features

Supports AmpliconSeq, WES, WGS

Uses proven open-source packages and frameworks


---

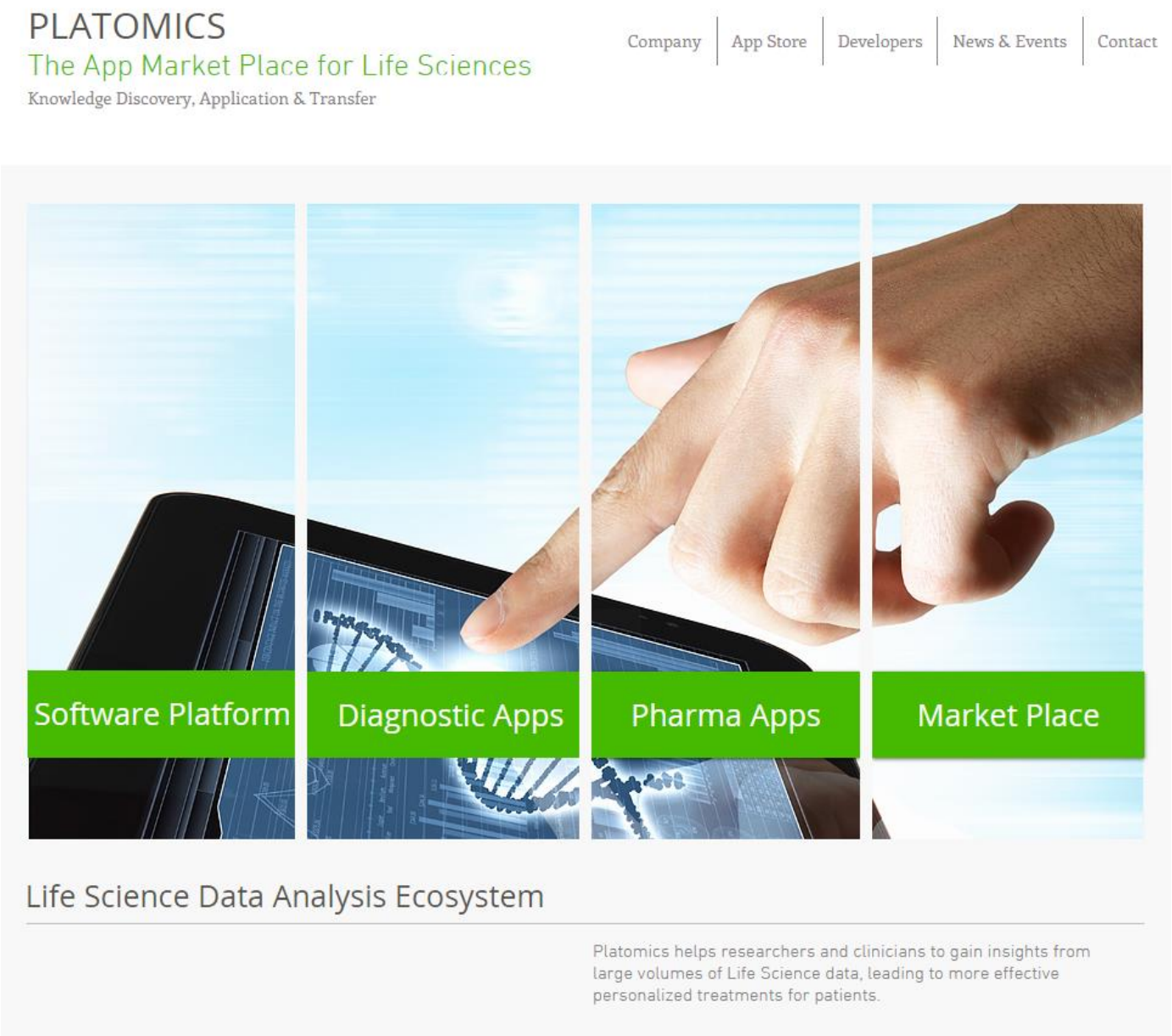
Transformation of variant coordinates into Transcript HGVS

Variant identification with multiple tools

→ Merging of variants from different callers

| | | | | | | |
|---|---|---|---|---|---|---|
| C(330) > T(29 | 47.45% | het | NM_000059.3:c.9038C>T | chr13:g.32953971C>T | non_syn_coding | F;S |
| A(952) > C(8 | 47.15% | het | NM_000059.3:c.*105A>C | chr13:g.32973012A>C | UTR_3_prime | F;S |
| C(1391) > T(9 | 40.57% | het | NM_007294.3:c.4956G>A | chr17:g.41222975C>T | non_syn_coding | F;S |

# Integration into Platomics Platform



# Integration into Platomics Platform

Web-based

Easy installation of server part

App-store concept

Customization

NGS Rare diseases Software

# Platomics

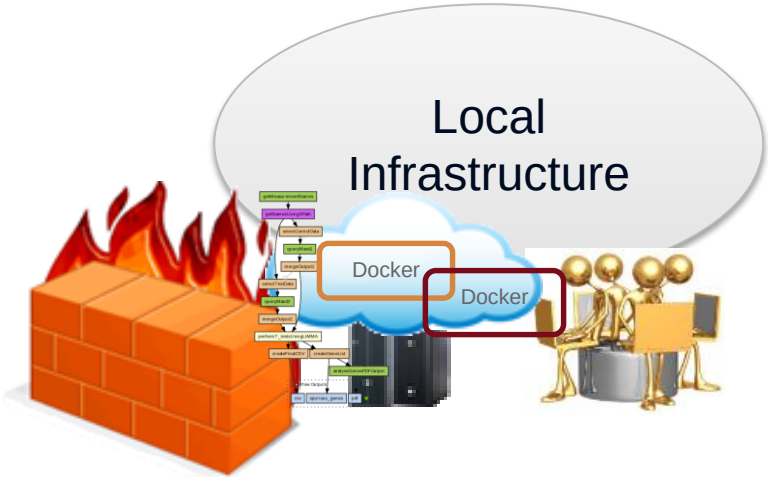
## Remote deployment

- Data access secured through user management
- Sharing of data

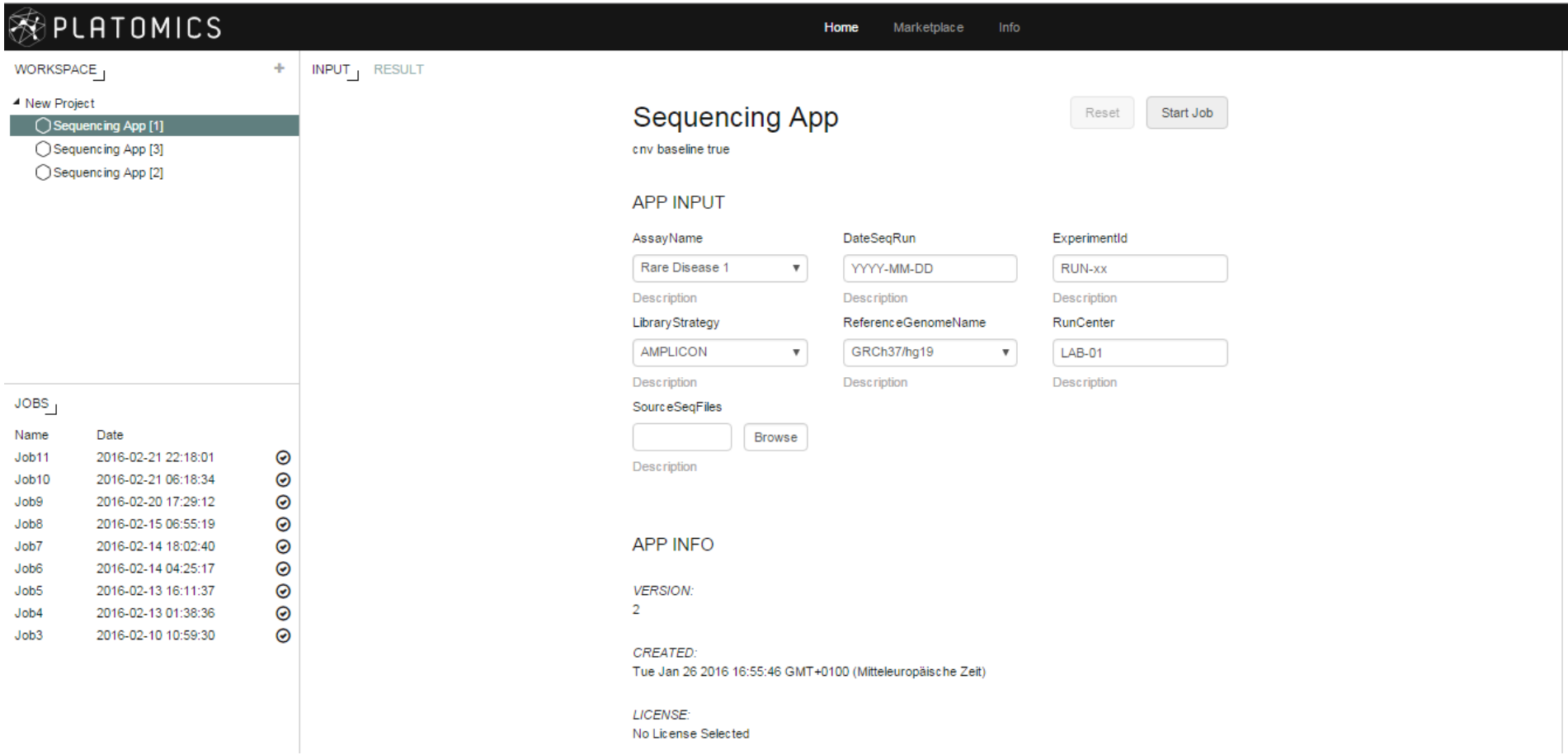

## Local deployment

- Only accessible through local network
- Data stored on local infrastructure

→ data security



# Integration into Platomics platform



# Results

INPUT RESULT

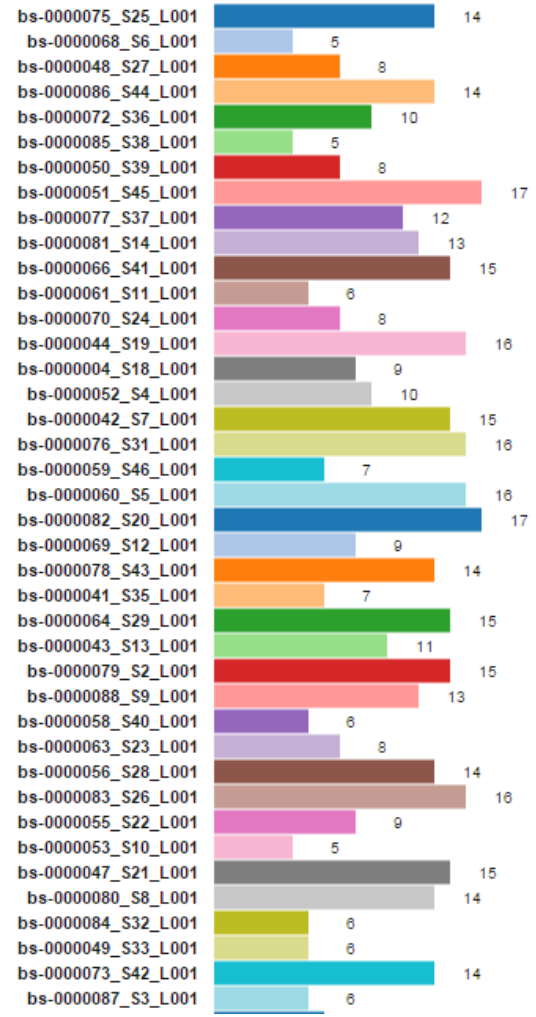
EXPERIMENT PATIENTS VARIANTS CNV GENES VALIDATION

Info

| | |
|---|---|
| **Experiment Id:** | RUN-12 |
| **Assay Name:** | BRCA1&2 CNV [as-3] |
| **Run Center:** | LAB-01 |
| **Date Seq Run:** | YYYY-MM-DD |
| **SeqPlatform:** | Illumina_MiSeq |
| **DateSeqAnalysis:** | 2016-02-14 |

Pathogenic serverity variants

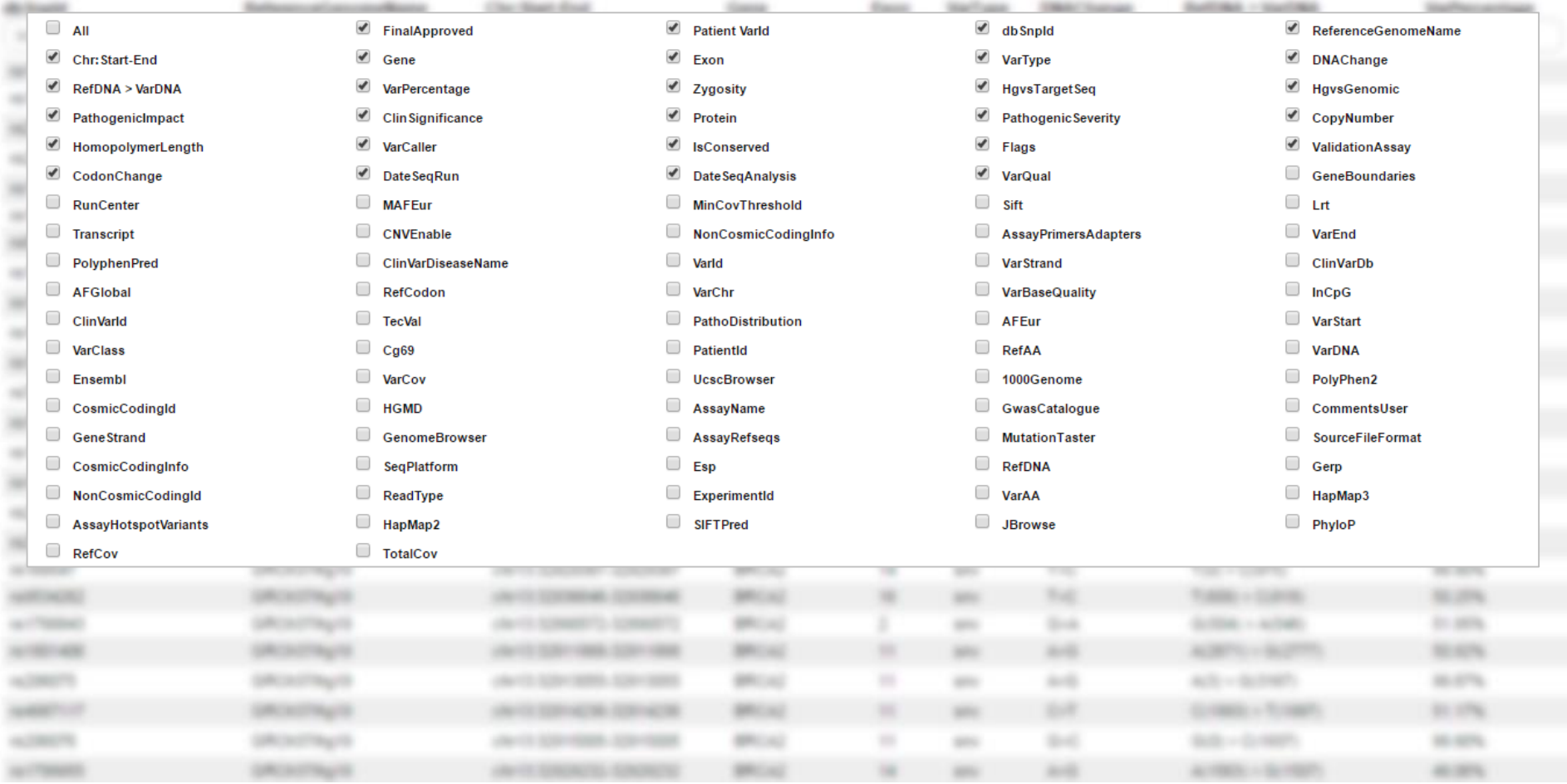

Variants per Patient



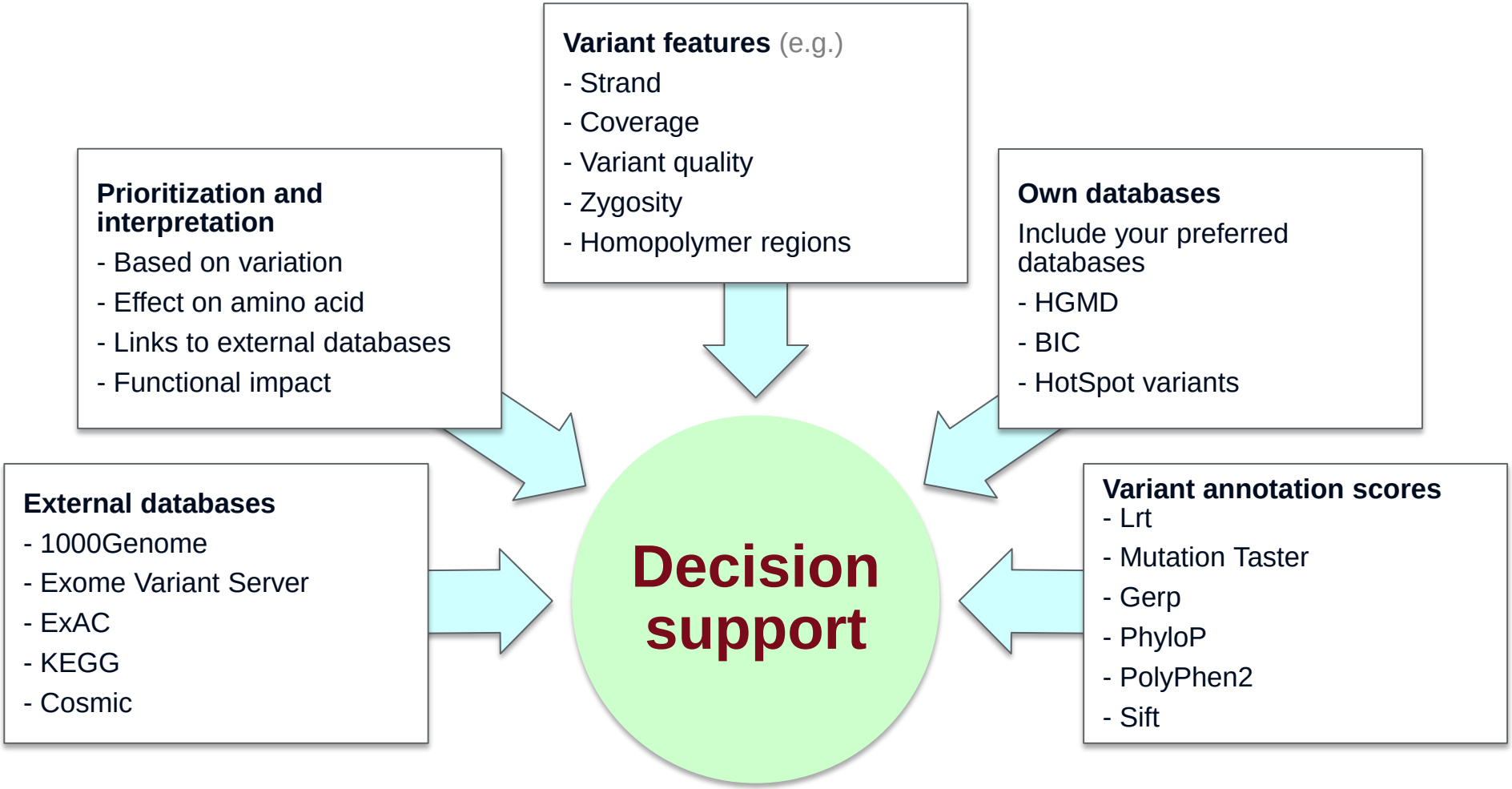
# Results

| dbSnpId | ReferenceGenomeName | Chr:Start-End | Gene | Exon | VarType | DNAChange | RefDNA > VarDNA | VarPercentage |
|---|---|---|---|---|---|---|---|---|
| search... | search... | search... | search... | searc | search.. | search... | search... | search... |
| rs1799943 | GRCh37/hg19 | chr13:32890572-32890572 | BRCA2 | 2 | snv | G>A | G(0) > A(1304) | 99.54% |
| rs1801406 | GRCh37/hg19 | chr13:32911888-32911888 | BRCA2 | 11 | snv | A>G | A(20) > G(7821) | 99.67% |
| rs206075 | GRCh37/hg19 | chr13:32913055-32913055 | BRCA2 | 11 | snv | A>G | A(2) > G(3868) | 99.92% |
| rs206076 | GRCh37/hg19 | chr13:32915005-32915005 | BRCA2 | 11 | snv | G>C | G(1) > C(2728) | 99.85% |
| rs1799955 | GRCh37/hg19 | chr13:32929232-32929232 | BRCA2 | 14 | snv | A>G | A(2066) > G(1974) | 48.79% |
| rs169547 | GRCh37/hg19 | chr13:32929387-32929387 | BRCA2 | 14 | snv | T>C | T(4) > C(1094) | 99.64% |
| rs9534262 | GRCh37/hg19 | chr13:32936646-32936646 | BRCA2 | 16 | snv | T>C | T(986) > C(966) | 49.46% |
| rs1799966 | GRCh37/hg19 | chr17:41223094-41223094 | BRCA1 | 16 | snv | T>C | T(0) > C(909) | 100.00% |
| rs1060915 | GRCh37/hg19 | chr17:41234470-41234470 | BRCA1 | 12 | snv | A>G | A(0) > G(1283) | 99.92% |
| rs16942 | GRCh37/hg19 | chr17:41244000-41244000 | BRCA1 | 10 | snv | T>C | T(15) > C(7544) | 99.67% |
| rs16941 | GRCh37/hg19 | chr17:41244435-41244435 | BRCA1 | 10 | snv | T>C | T(15) > C(6083) | 99.75% |
| rs799917 | GRCh37/hg19 | chr17:41244936-41244936 | BRCA1 | 10 | snv | G>A | G(14) > A(3740) | 99.44% |
| rs16940 | GRCh37/hg19 | chr17:41245237-41245237 | BRCA1 | 10 | snv | A>G | A(17) > G(9433) | 99.76% |
| rs1799949 | GRCh37/hg19 | chr17:41245466-41245466 | BRCA1 | 10 | snv | G>A | G(26) > A(4910) | 98.38% |
| rs1801406 | GRCh37/hg19 | chr13:32911888-32911888 | BRCA2 | 11 | snv | A>G | A(2945) > G(3037) | 50.67% |
| rs206075 | GRCh37/hg19 | chr13:32913055-32913055 | BRCA2 | 11 | snv | A>G | A(8) > G(3069) | 99.71% |
| rs206076 | GRCh37/hg19 | chr13:32915005-32915005 | BRCA2 | 11 | snv | G>C | G(0) > C(1963) | 99.95% |
| rs169547 | GRCh37/hg19 | chr13:32929387-32929387 | BRCA2 | 14 | snv | T>C | T(0) > C(975) | 99.90% |
| rs9534262 | GRCh37/hg19 | chr13:32936646-32936646 | BRCA2 | 16 | snv | T>C | T(809) > C(818) | 50.25% |
| rs1799943 | GRCh37/hg19 | chr13:32890572-32890572 | BRCA2 | 2 | snv | G>A | G(504) > A(546) | 51.95% |
| rs1801406 | GRCh37/hg19 | chr13:32911888-32911888 | BRCA2 | 11 | snv | A>G | A(2671) > G(2777) | 50.92% |
| rs206075 | GRCh37/hg19 | chr13:32913055-32913055 | BRCA2 | 11 | snv | A>G | A(3) > G(3167) | 99.87% |
| rs4987117 | GRCh37/hg19 | chr13:32914236-32914236 | BRCA2 | 11 | snv | C>T | C(1893) > T(1997) | 51.17% |
| rs206076 | GRCh37/hg19 | chr13:32915005-32915005 | BRCA2 | 11 | snv | G>C | G(0) > C(1937) | 99.90% |
| rs1799955 | GRCh37/hg19 | chr13:32929232-32929232 | BRCA2 | 14 | snv | A>G | A(1563) > G(1507) | 49.06% |

# Results – fully customizable



# Module 3 – Annotation & Prioritization

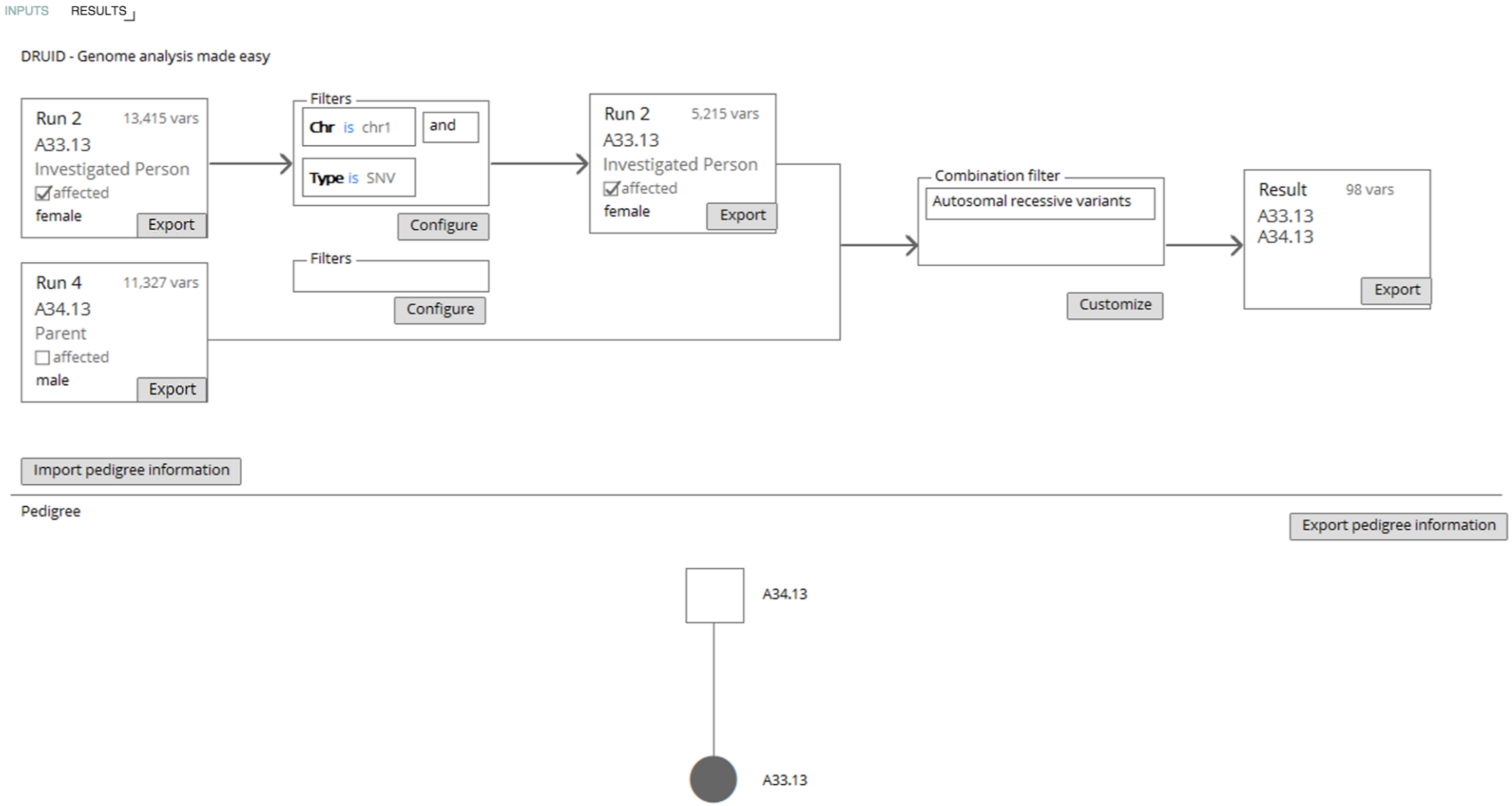


# Results

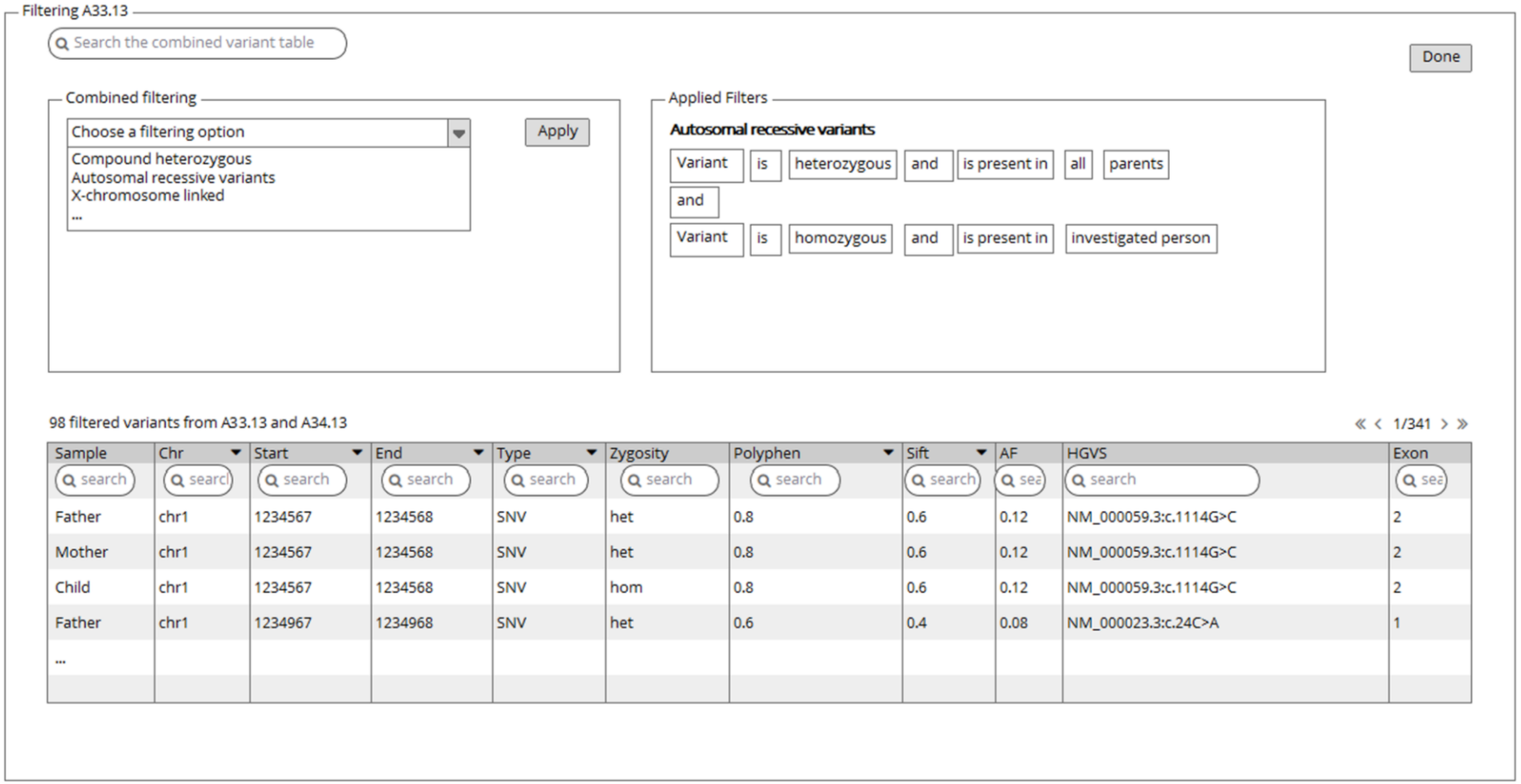
| dbSnpId | ReferenceGenomeName | Chr:Start-End | Gene | Exon | VarType | DNAChange | RefDNA > VarDNA | VarPercentage |
|---|---|---|---|---|---|---|---|---|
| search... | search... | search... | search... | searc | search... | search... | search... | search... |
| rs1799943 | GRCh37/hg19 | chr13:32890572-32890572 | BRCA2 | 2 | snv | G>A | G(0) > A(1304) | 99.54% |
| rs1801406 | GRCh37/hg19 | chr13:32911888-32911888 | BRCA2 | 11 | snv | A>G | A(20) > G(7821) | 99.67% |
| rs206075 | GRCh37/hg19 | chr13:32913055-32913055 | BRCA2 | 11 | snv | A>G | A(2) > G(3868) | 99.92% |
| rs206076 | GRCh37/hg19 | | | | | | G(1) > C(2728) | 99.85% |
| rs1799955 | GRCh37/hg19 | | | | | | A(2066) > G(1974) | 48.79% |
| rs169547 | GRCh37/hg19 | | | | | | T(4) > C(1094) | 99.64% |
| rs9534262 | GRCh37/hg19 | | | | | | T(966) > C(966) | 49.46% |
| rs1799966 | GRCh37/hg19 | chr17:41223094-41223094 | BRCA1 | 16 | snv | T>C | T(0) > C(909) | 100.00% |
| rs1060915 | GRCh37/hg19 | chr17:41234470-41234470 | BRCA1 | 12 | snv | A>G | A(0) > G(1283) | 99.92% |
| rs16942 | GRCh37/hg19 | chr17:41244000-41244000 | BRCA1 | 10 | snv | T>C | T(15) > C(7544) | 99.67% |
| rs16941 | GRCh37/hg19 | chr17:41244435-41244435 | BRCA1 | 10 | snv | T>C | T(15) > C(6083) | 99.75% |
| rs799917 | GRCh37/hg19 | chr17:41244936-41244936 | BRCA1 | 10 | snv | G>A | G(14) > A(3740) | 99.44% |
| rs16940 | GRCh37/hg19 | chr17:41245237-41245237 | BRCA1 | 10 | snv | A>G | A(17) > G(9433) | 99.76% |
| rs1799949 | GRCh37/hg19 | chr17:41245466-41245466 | BRCA1 | 10 | snv | G>A | G(26) > A(4910) | 98.38% |
| rs1801406 | GRCh37/hg19 | chr13:32911888-32911888 | BRCA2 | 11 | snv | A>G | A(2945) > G(3037) | 50.67% |
| rs206075 | GRCh37/hg19 | chr13:32913055-32913055 | BRCA2 | 11 | snv | A>G | A(8) > G(3069) | 99.71% |
| rs206076 | GRCh37/hg19 | chr13:32915005-32915005 | BRCA2 | 11 | snv | G>C | G(0) > C(1963) | 99.95% |
| rs169547 | GRCh37/hg19 | chr13:32929387-32929387 | BRCA2 | 14 | snv | T>C | T(0) > C(975) | 99.90% |
| rs9534262 | GRCh37/hg19 | chr13:32936646-32936646 | BRCA2 | 16 | snv | T>C | T(809) > C(818) | 50.25% |
| rs1799943 | GRCh37/hg19 | chr13:32890572-32890572 | BRCA2 | 2 | snv | G>A | G(504) > A(546) | 51.95% |
| rs1801406 | GRCh37/hg19 | chr13:32911888-32911888 | BRCA2 | 11 | snv | A>G | A(2871) > G(2777) | 50.92% |
| rs206075 | GRCh37/hg19 | chr13:32913055-32913055 | BRCA2 | 11 | snv | A>G | A(3) > G(3167) | 99.87% |
| rs4987117 | GRCh37/hg19 | chr13:32914236-32914236 | BRCA2 | 11 | snv | C>T | C(1893) > T(1997) | 51.17% |
| rs206076 | GRCh37/hg19 | chr13:32915005-32915005 | BRCA2 | 11 | snv | G>C | G(0) > C(1937) | 99.90% |
| rs1799955 | GRCh37/hg19 | chr13:32929232-32929232 | BRCA2 | 14 | snv | A>G | A(1563) > G(1507) | 49.06% |

… still, a long list of variants

# Filtering


INPUTS
RESULTS
DRUID - Genome analysis made easy
Run 2
13,415 vars
A33.13
Investigated Person
affected
female
Export
Run 4
11,327 vars
A34.13
Parent
affected
male
Export
Filters
Chr is chr1
and
Type is SNV
Configure
Filters
Configure
Run 2
5,215 vars
A33.13
Investigated Person
affected
female
Export
Combination filter
Autosomal recessive variants
Customize
Result
98 vars
A33.13
A34.13
Export
Import pedigree information
Pedigree
Export pedigree information
A34.13
A33.13

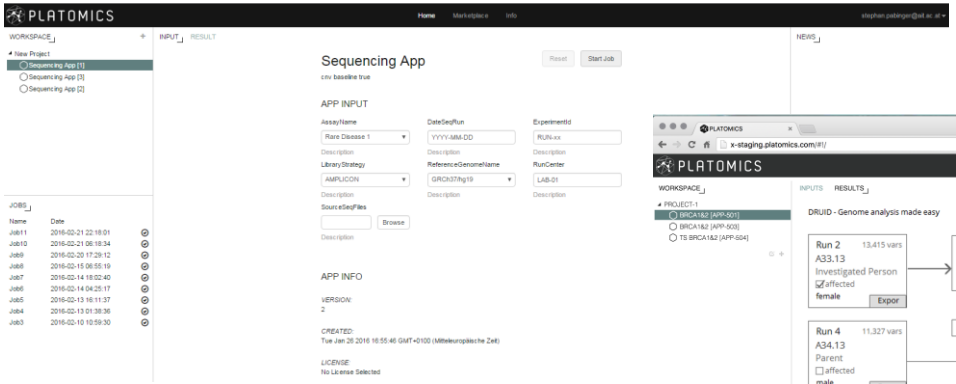
# Interactive filtering and prioritization of variants

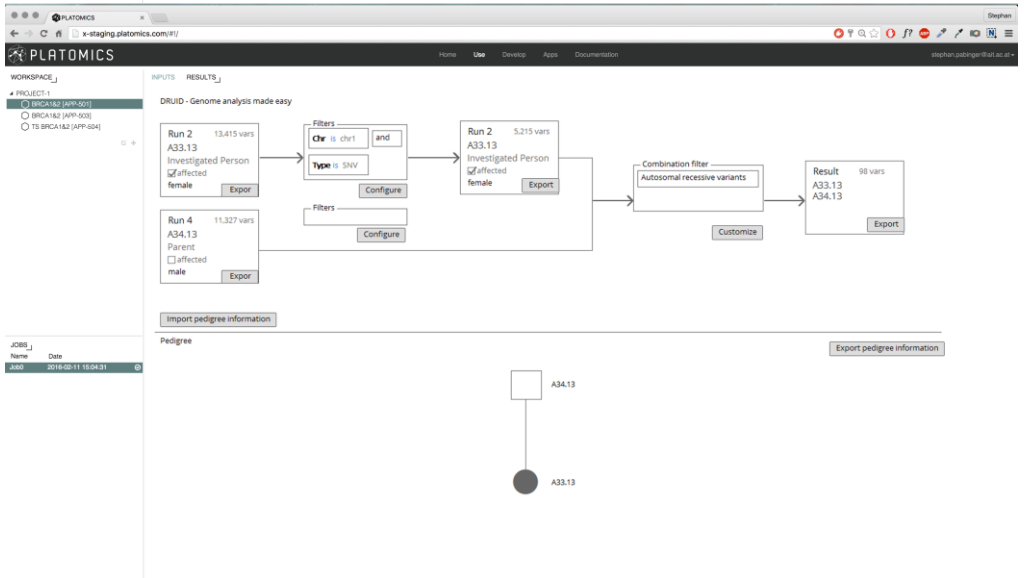



| Sample | Chr | Start | End | Type | Zygosity | Polyphen | Sift | AF | HGVS | Exon |
|---|---|---|---|---|---|---|---|---|---|---|
| Father | chr1 | 1234567 | 1234568 | SNV | het | 0.8 | 0.6 | 0.12 | NM_000059.3:c.1114G>C | 2 |
| Mother | chr1 | 1234567 | 1234568 | SNV | het | 0.8 | 0.6 | 0.12 | NM_000059.3:c.1114G>C | 2 |
| Child | chr1 | 1234567 | 1234568 | SNV | hom | 0.8 | 0.6 | 0.12 | NM_000059.3:c.1114G>C | 2 |
| Father | chr1 | 1234967 | 1234968 | SNV | het | 0.6 | 0.4 | 0.08 | NM_000023.3:c.24C>A | 1 |
| ... | | | | | | | | | | |

# Summary

- Software for variant identification and annotation
- Integration into a web-based system (Platomics)
- Intuitive filtering mechanism



# Acknowledgments



www.ait.ac.at

- Klemens Vierlinger


www.cemm.at

- Ana Krolo
- Tatjana T. Hirschmugl
- Kaan Boztug
- Christoph Bock


www.platomics.com

- Denis Katic
- Martin Dulovits
- Gregor Rosenauer
- Albert Kriegner

# Collaboration oportunities

**Horizon 2020**

PM03 - Diagnostic characterisation of rare diseases

PM08 - New therapies for rare diseases


Stephan Pabinger

AIT Austrian Institute of Technology

Bioinformatics

Molecular Diagnostics

stephan.pabinger@ait.ac.at

+43 50550 4409