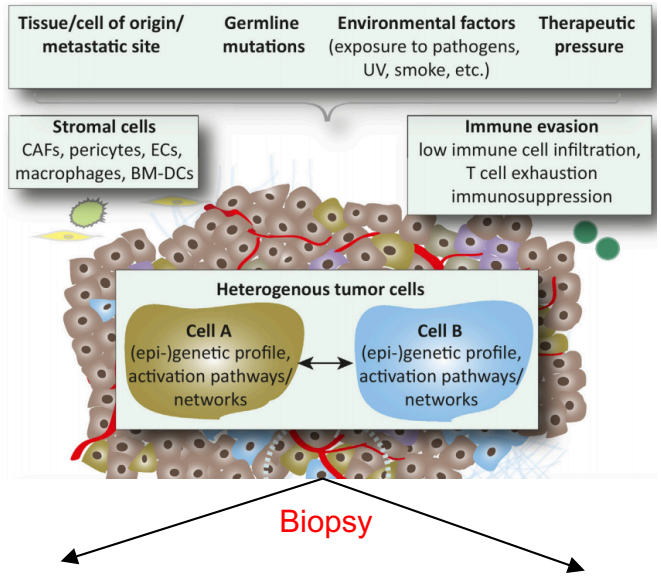
Use case oncology

Personal Cancer Genome Reporter (PCGR):
variant interpretation report for precision oncology

Dr. Sigve Nakken
Norwegian Cancer Genomics Consortium (NCGC)
Eivind Hovig group, Dept. of Tumor Biology
Institute for Cancer Research
Oslo, Norway

Omics for precision oncology – overview (I)



1. Determinants of tumor pathogenesis

2. Molecular profiling – genomics & transcriptomics

DNA-seq (exome)

- Single nucleotide variants
- Insertions/deletions
- Copy number aberrations
- (Rearrangements)

RNA-seq (exome)

- Relative gene expression
- Gene fusions
- Pathway activity

Illustration adapted from Senft et al., Trends Mol Med, 2017

Omics for precision oncology – overview (II)

DNA-seq (exome)
- Single nucleotide variants
- Insertions/deletions
- Copy number aberrations
- (Rearrangements)

RNA-seq (exome)
- Relative gene expression
- Gene fusions
- Pathway activity

2. Molecular profiling – genomics & transcriptomics

- *Oncogene/tumor suppressor gene mutations*
- *Oncogene activity*
- *Cell type composition status*
- *Tumor heterogeneity/clonality estimation*
- *Tissue-of-origin prediction*
- *Contribution of mutational signatures*
- *Microsatellite instability status*
- *BRCAness status*
- *Mutational burden*
- *Clinical trial matcher*

3. Measures to inform precision therapy

Omics for precision oncology – overview (II)

Bioinformatics: Robust sequencing pipelines

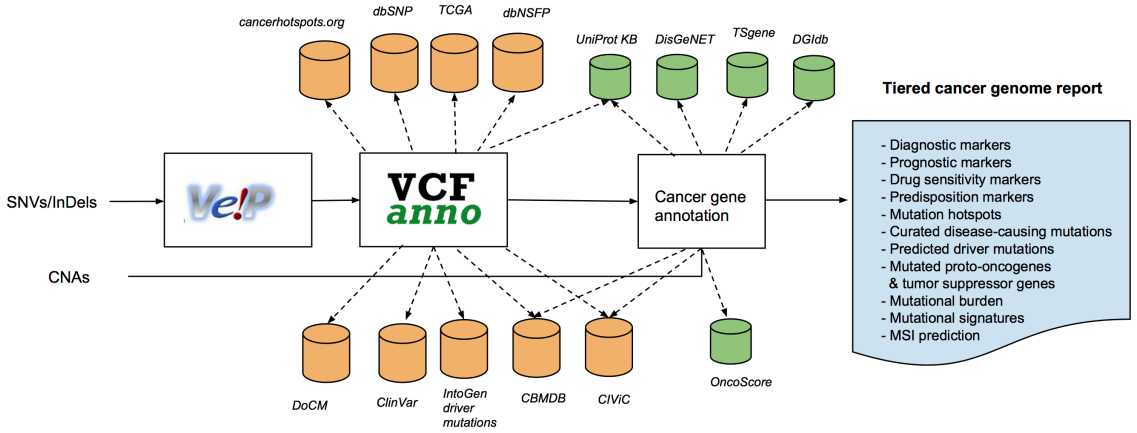
2. Molecular profiling – genomics & transcriptomics

Bioinformatics: Translation of molecular profiling data

PCGR

3. Measures to inform precision therapy

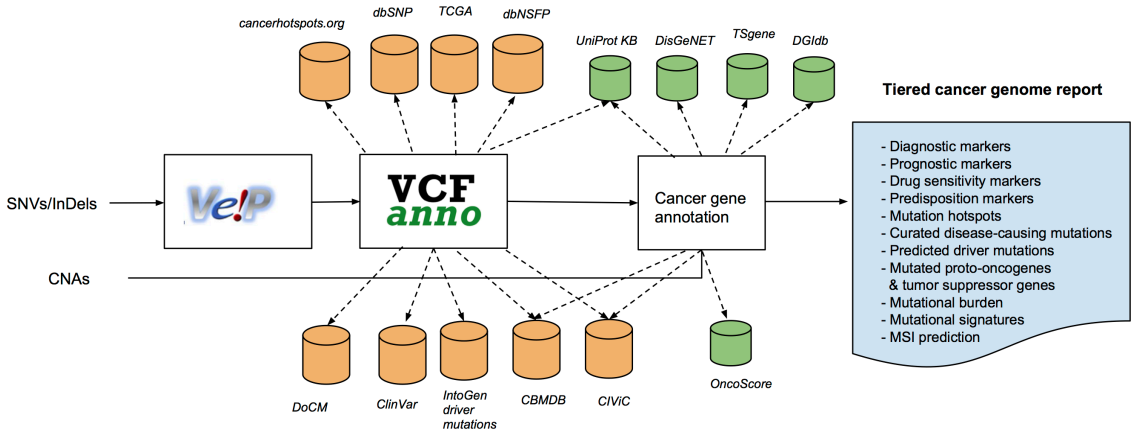
Personal Cancer Genome Reporter (PCGR)



Nakken et al. (accepted, Bioinformatics)

- What is it?
  - Technical architecture
    - Knowledge resources
  - Input data
  - Output data
  - Future plans

PCGR: architecture



- Stand-alone software workflow
- Freely available, open source
- Built using Docker technology
- Users need to download
  - The software (Docker image and script to run)
  - Data bundle
    - contains up-to-date datasets for functional annotation
- *https://github.com/sigven/pcgr*

PCGR: knowledge resources

- Targeted cancer drugs
- Known biomarkers for prognosis/diagnosis/predisposition
- Known biomarkers for drug sensitivity/resistance
- Variant frequency across tumor types
- Curated gene-cancer associations
- Mutational hotspots
- Predicted driver mutations
- Signaling pathways
- Proto-oncogenes/tumor suppressors
- Prediction of variant effect on protein function

PCGR input: DNA variation types

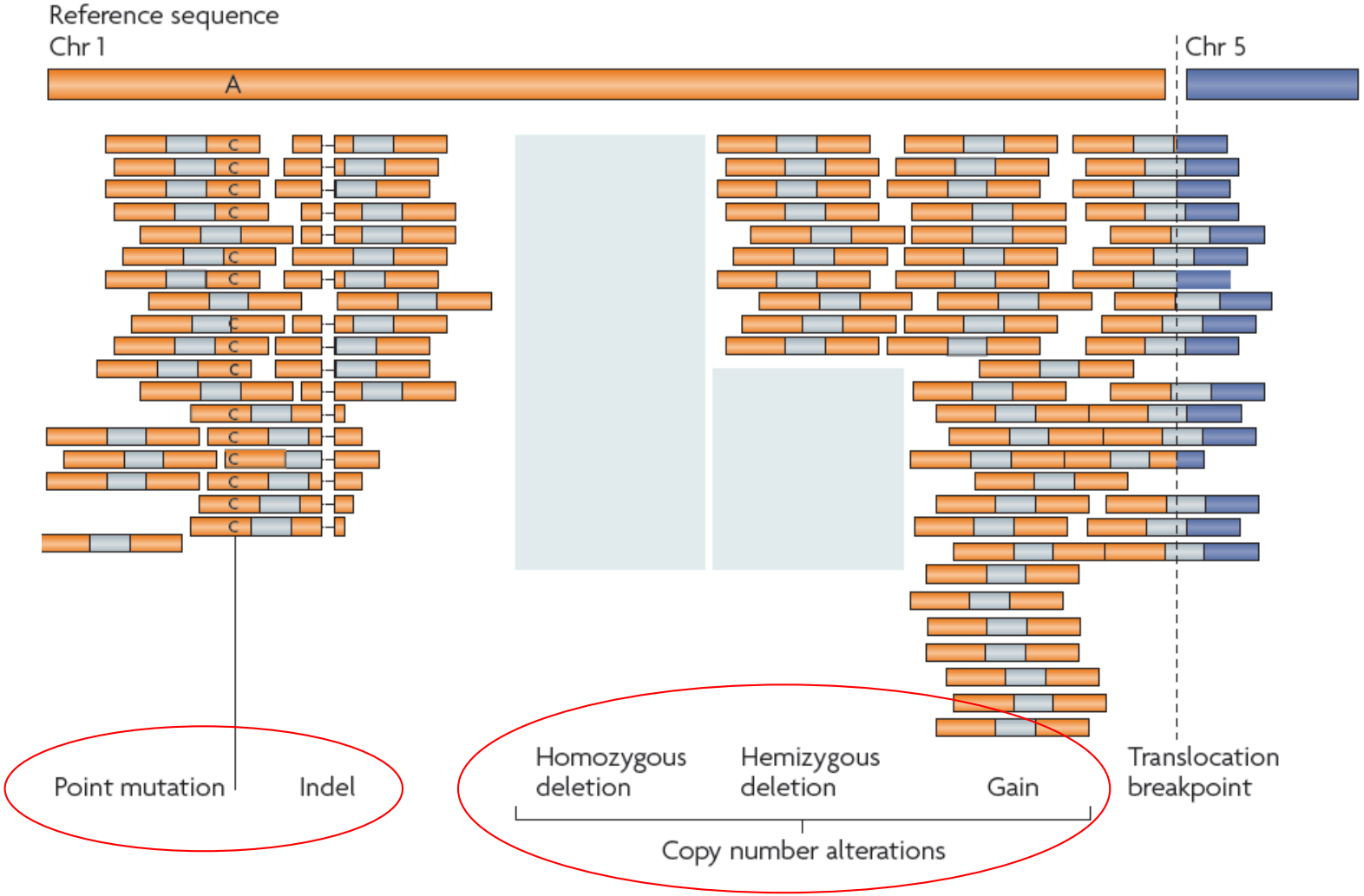
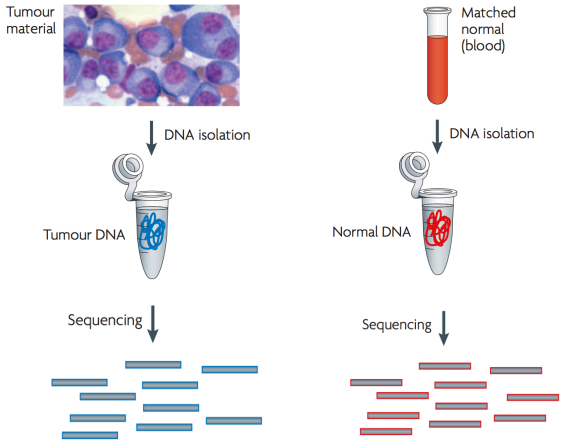
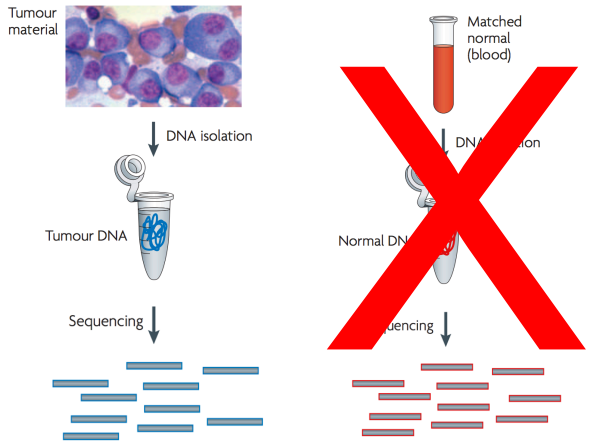


Illustration from Meyerson et al. Nat Rev Genet 2010

PCGR input: tumor-normal vs. tumor-only



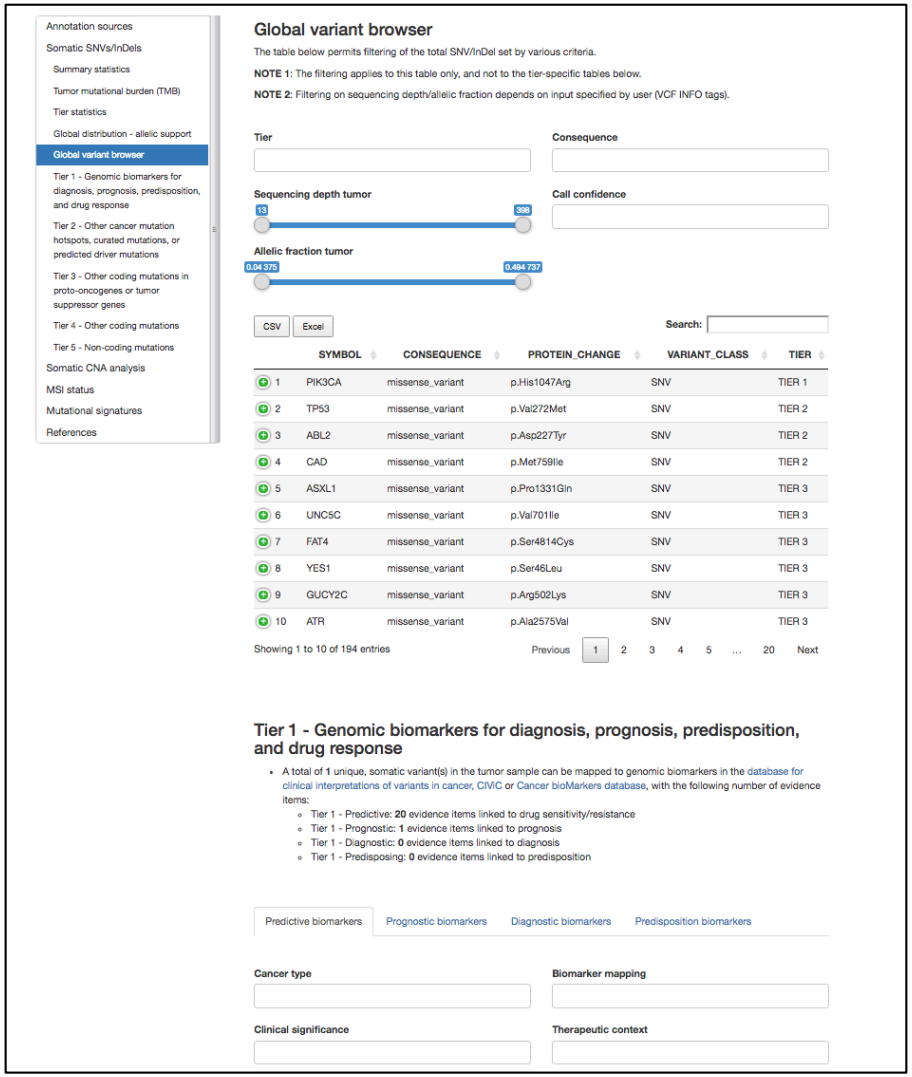
- Variant calling
  - Somatic variants
- Most robust/precise approach



- Variant calling
  - Somatic variants
  - Germline variants
- Enrichment of somatic events by exclusion of known germline variants
  - gnomAD, 1000 Genome Project etc.
- Challenges:
  1. Each individual is estimated to carry an extensive set of rare variants (i.e. *singletons*)
  2. Ethnic subpopulations are under-represented in public germline variant databases

PCGR: output

- Interactive, tiered-structured genome report
  - Adopting proposed recommendations/ nomenclature (oncology)
  - Links to underlying sources (biomedical literature, trials etc.)
- Interprets SNVs/InDels/CNAs in the context of known actionable variants
- Mutational burden
- Infers contribution of known mutational signatures
- Microsatellite stability predictor
- Supports tumor-only runs
- ++



PCGR output: mutational burden

- Genomic biomarker for immunotherapy response
- Provided by user
  - Target size
- Calculates number of mutations pr. Mb in target
- For samples with high underlying TMB, sequencing targeted panels is sufficient

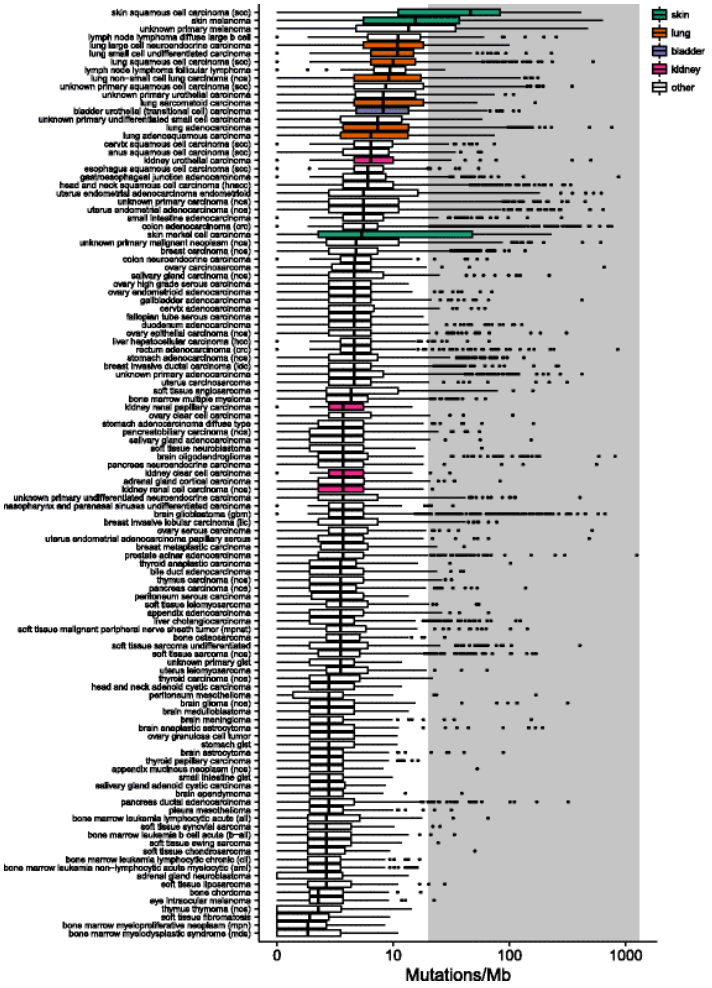


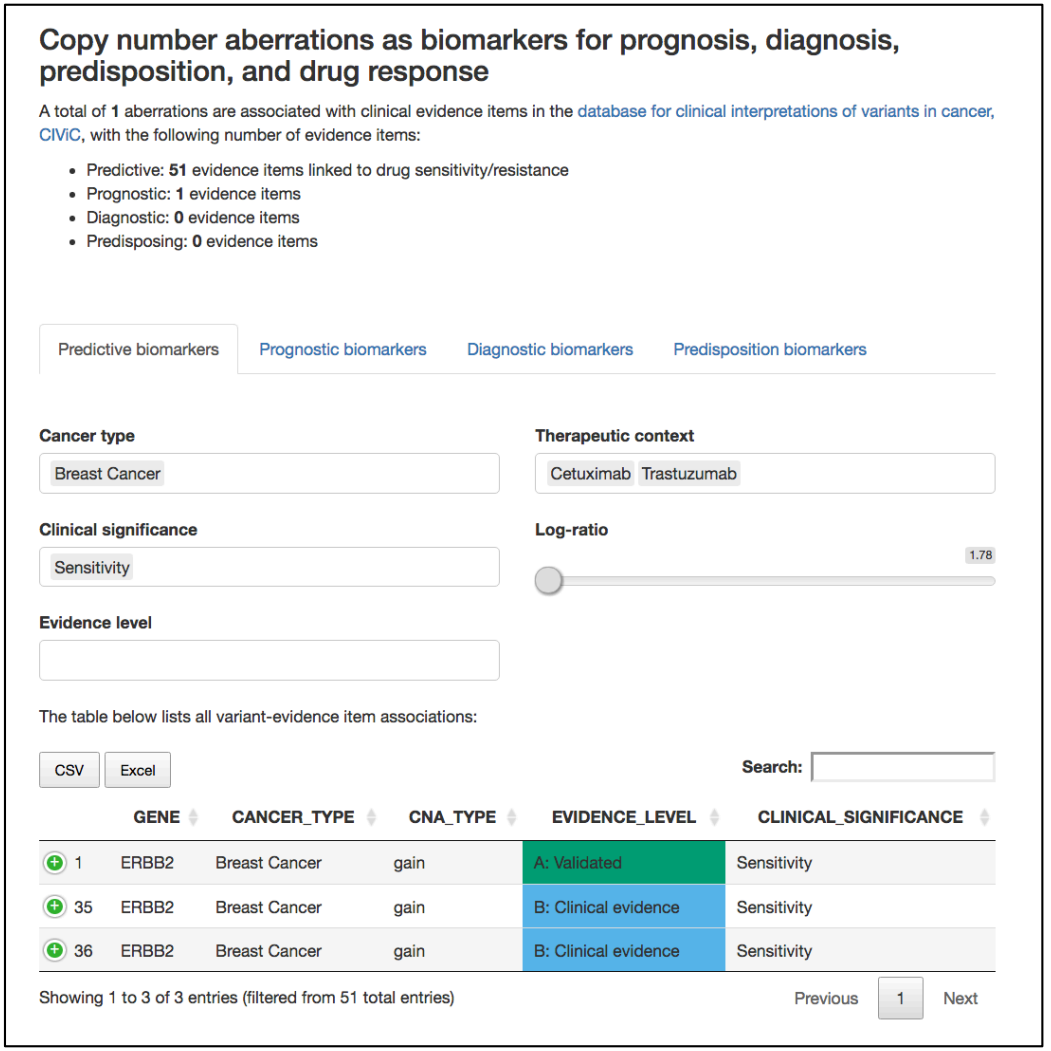
Figure from Chalmers et al. Genome Med, 2017

PCGR output: tier structure

- SNVs/InDels
  - Tier 1: Actionable variants
  - Tier 2: Mutational hotspots, predicted driver mutations, or known disease-causing mutations
  - Tier 3: Other cancer gene mutations
  - Tier 4: Other coding mutations
  - Tier 5: Non-coding mutations
- CNAs
  - Amplifications and homozygous deletions (as defined by the user)
  - Biomarkers and drug targets

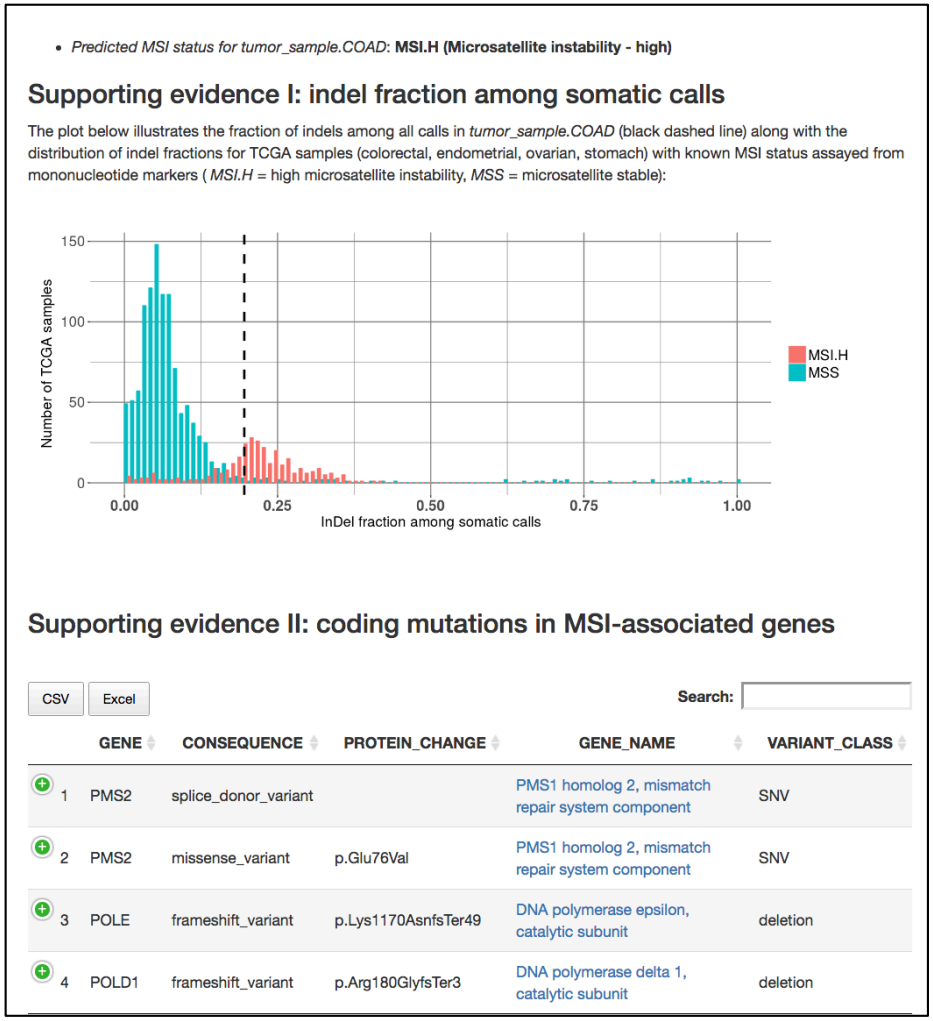
PCGR output: clinical evidence items

- Each actionable variant associated with one or more evidence item
- A piece of information that has been manually curated from trustable medical literature about a variant or genomic 'event' that has implications in cancer predisposition, diagnosis (*aka* molecular classification), prognosis, or predictive response to therapy
  - Therapeutic context
  - Evidence level
  - Tumor type
  - Evidence type (Prognostic, Predictive etc.)
  - Etc.



PCGR output: Microsatellite instability status

- Prediction of MSI from SNV/InDel distribution
  - A statistical classifier trained using ~1500 samples with known MSI status assayed from mononucleotide markers
  - ***MSI.H*** vs. ***MSS***
  - Discriminatory features
    - Fraction of indels
    - Fraction of indels in repeat sequences
    - Loss-of-function mutations in MMR genes
  - Sensitivity: 0.90
  - Specificity: 0.997



- *Predicted MSI status for tumor_sample.COAD*: **MSI.H (Microsatellite instability - high)**

Supporting evidence I: indel fraction among somatic calls

The plot below illustrates the fraction of indels among all calls in *tumor_sample.COAD* (black dashed line) along with the distribution of indel fractions for TCGA samples (colorectal, endometrial, ovarian, stomach) with known MSI status assayed from mononucleotide markers (*MSI.H* = high microsatellite instability, *MSS* = microsatellite stable):

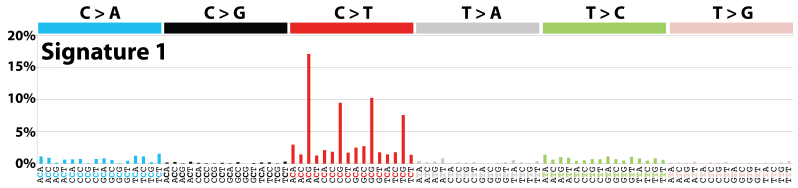
Supporting evidence II: coding mutations in MSI-associated genes

CSV Excel Search:

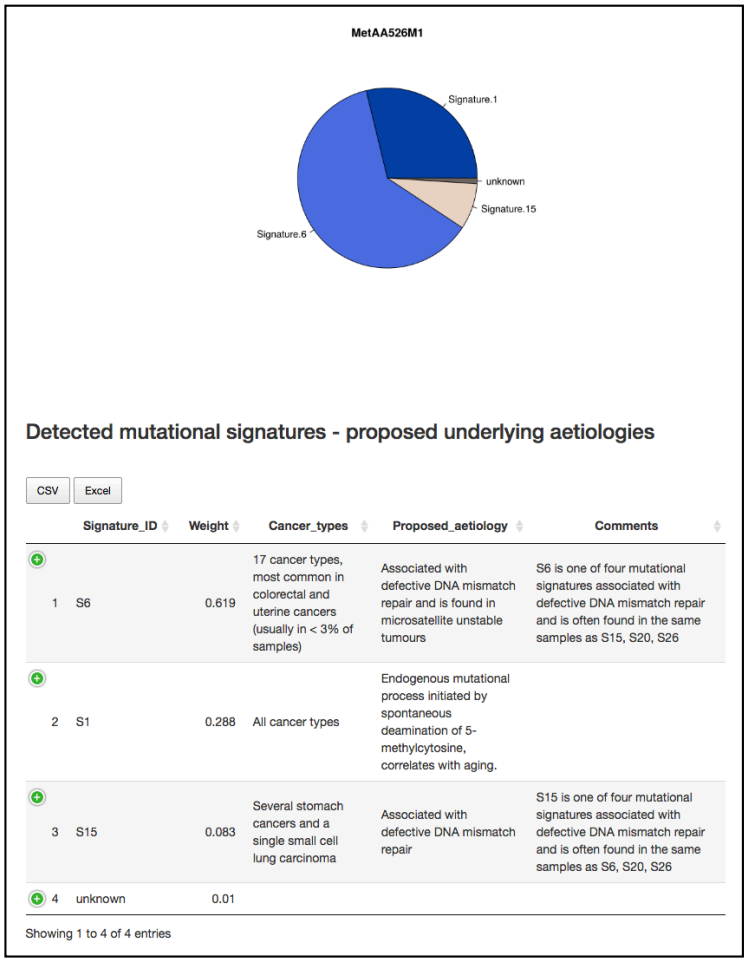
	GENE	CONSEQUENCE	PROTEIN_CHANGE	GENE_NAME	VARIANT_CLASS
1	PMS2	splice_donor_variant		PMS1 homolog 2, mismatch repair system component	SNV
2	PMS2	missense_variant	p.Glu76Val	PMS1 homolog 2, mismatch repair system component	SNV
3	POLE	frameshift_variant	p.Lys1170AsnfsTer49	DNA polymerase epsilon, catalytic subunit	deletion
4	POLD1	frameshift_variant	p.Arg180GlyfsTer3	DNA polymerase delta 1, catalytic subunit	deletion

PCGR output: mutational signatures

- Mutational signatures - the pattern of somatic mutations imprinted by the activity of a mutational process
- What is meant by “pattern”?
  - Distinct distribution of context-dependent mutations



- PCGR – reports contribution of *known* mutational signatures in a single tumor
- Emerging evidence for clinical relevance of specific signatures
  - *BRCAness*



PCGR: future plans

- New functionality
  - Add support for additional data-types (i.e. RNA-seq)
    - Opens up possibilities for a whole new range of analyses
      - Estimation of cell type composition (immune infiltration)
      - Infer pathway activity
      - ++
  - Customize results according to the query tumor
- Flexibility
  - Enable developers to contribute predictions/analysis modules (plug-ins)
    - Feedback from users/developers
  - Customized report template

PCGR: possible sharing strategies

- Elixir
  - bio.tools registry

- A web portal for reproducible report generation
  - ***wPCGR***
  - More accessible to users with non-computational background

Acknowledgements

- NCGC PI's
- NCGC Bioinformatics group
- Genomics Core Facility, Oslo University Hospital
- Eivind Hovig

Thank you for your attention!

github.com/sigven/pcgr