Variation Annotation Tool (VAT)

mtg-anno

01/11/2011

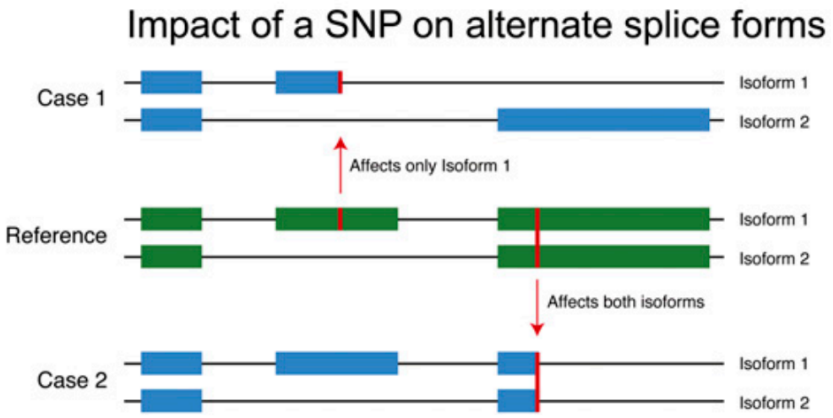
Lukas Habegger

Objective

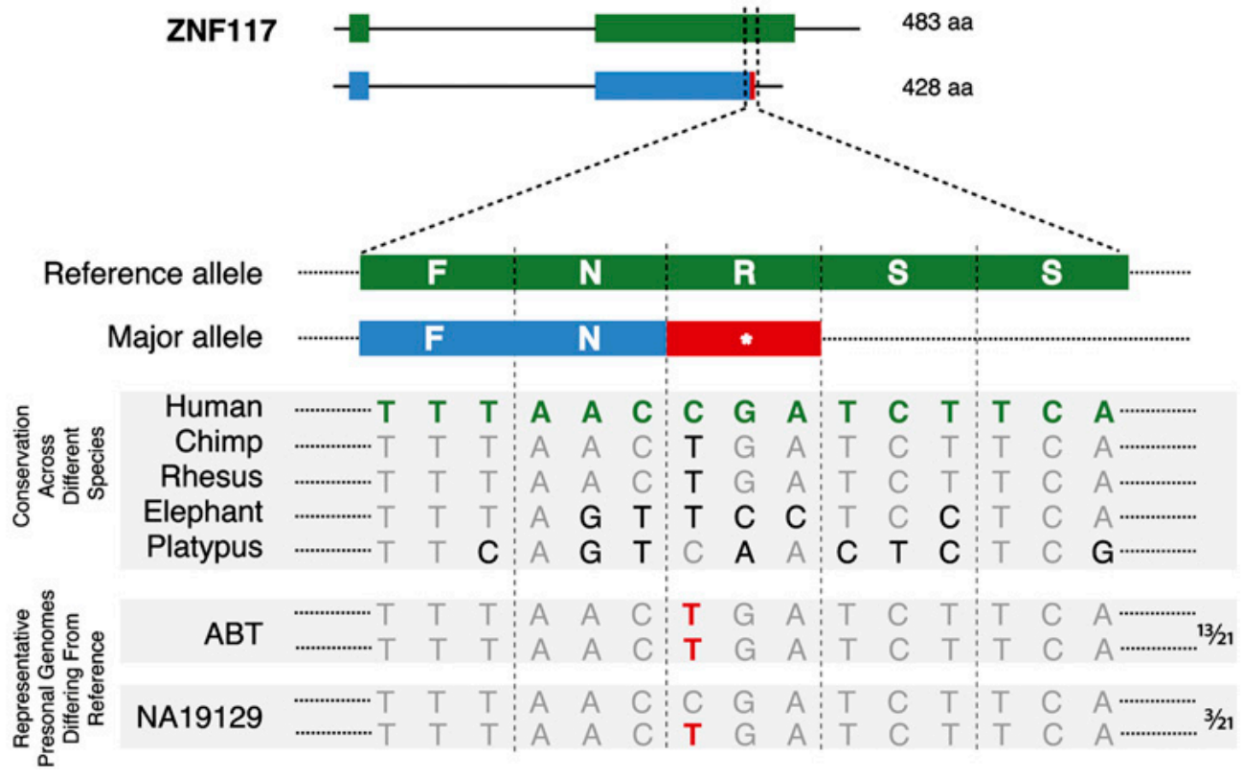
- To annotate genetic variants from personal genomes
  - SNPs
  - Indels
- Efficient algorithm
  - Command line
  - Web-interface
- Visualize the results

Types of variants

- SNPs
  - Synonymous
  - Non-synonymous
  - Premature stop
  - Removed stop
  - Splice overlap
- Indels
  - Insertions (frameshift, non-frameshift)
  - Deletions (frameshift, non-frameshift)
  - Splice overlap
  - Complex

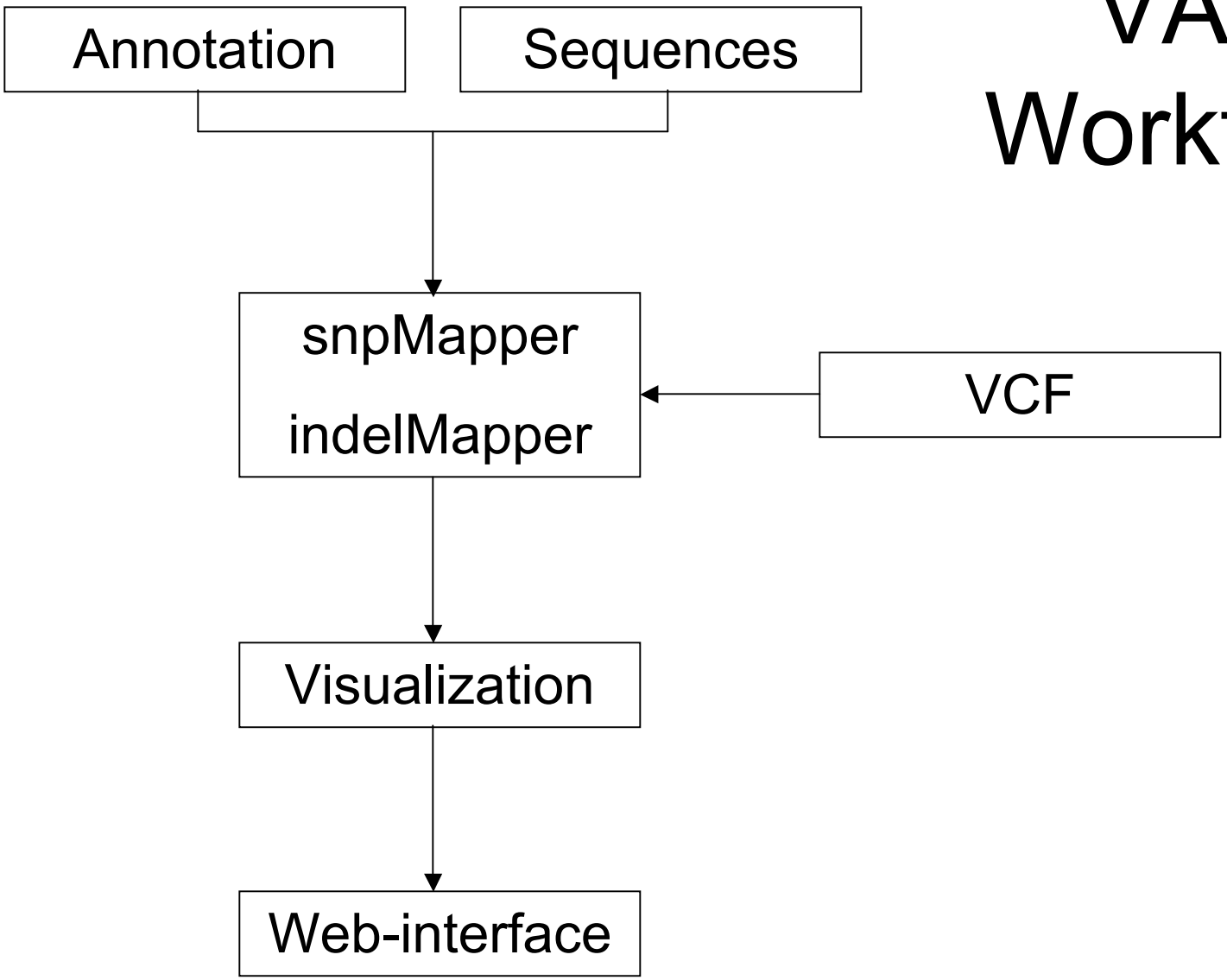


Manifestation of genetic variants



Balasubramanian et al., *Genes Dev.*, 2011

VAT Workflow



Implementation and Performance

- All programs are implemented in C
  - VCF (Variant call format) module
  - VAF (Variant annotation format) module

- Visualization
  - GD library

- Web-interface
  - jQuery, JSON, Ajax

- Performance (approximately **five times** faster than ANNOVAR)
  - Annotation of 10.5 SNPs (snpMapper)
    - Runtime: 100 seconds
    - Memory footprint: 150 MB

Web-interface

- Low-coverage
  - CEU
  - YRI
  - CHBJPT
- Trios
  - CEU
  - YRI

Results for data set: CEU.low_coverage

Show 25 entries Search:

Gene ID	Gene name	Number of transcripts	Number of synonymous SNPs	Number of nonsynonymous SNPs	Number of prematureStop SNPs	Number of removedStop SNPs	Number of spliceOverlaps	Number of insertions	Number of deletions	Details
ENSG00000000419	DPM1	6	0	1	0	0	0	0	0	Link
ENSG00000000457	SCYL3	4	5	2	0	0	0	0	0	Link
ENSG00000000460	C1orf112	4	2	3	0	0	0	0	0	Link
ENSG00000000938	FGR	5	1	0	0	0	0	0	0	Link
ENSG00000000971	CFH	5	4	5	0	0	0	0	0	Link
ENSG00000001036	FUCA2	5	1	5	0	0	0	0	0	Link
ENSG00000001084	GCLC	1	1	1	0	0	0	0	0	Link
ENSG00000001460	C1orf201	10	4	6	0	0	0	0	0	Link
ENSG00000001461	NIPAL3	9	1	1	0	0	0	0	0	Link
ENSG00000001561	ENPP4	2	1	1	0	0	0	0	0	Link
ENSG00000001626	CFTR	5	7	4	0	0	0	0	0	Link
ENSG00000001629	ANKIB1	2	1	1	0	0	0	0	0	Link
ENSG00000001630	CYP51A1	5	1	1	0	0	0	0	0	Link
ENSG00000001631	KRIT1	20	2	0	0	0	0	0	0	Link
ENSG00000002016	RAD52	4	0	1	0	0	0	1	0	Link
ENSG00000002330	BAD	3	1	1	0	0	0	0	0	Link
ENSG00000002726	ABP1	9	6	5	0	0	1	0	0	Link
ENSG00000002745	WNT16	3	0	3	0	0	0	2	0	Link
ENSG00000002746	HECW1	5	9	4	0	0	0	0	0	Link
ENSG00000002822	MAD1L1	15	5	2	0	0	0	0	0	Link
ENSG00000002834	LASP1	6	5	2	0	0	0	1	0	Link
ENSG00000002933	TMEM176A	5	2	6	0	0	0	0	0	Link
ENSG00000003056	M6PR	1	0	0	0	0	0	1	0	Link
ENSG00000003137	CYP26B1	2	1	1	0	0	0	0	0	Link
ENSG00000003147	ICA1	13	0	2	0	0	0	0	0	Link

Showing 1 to 25 of 15,853 entries

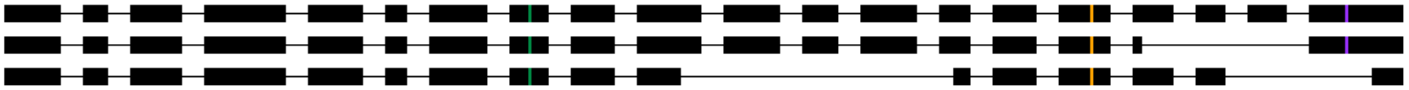
First Previous 1 2 3 4 5 Next Last

Results for ENSG00000068976 in data set CEU.low_coverage

Gene summary: PYGM (ENSG00000068976)

Transcript name	Transcript ID	Chromosome	Strand	Start	End	Number of exons	Transcript length
PYGM-001	ENST00000164139	chr11	-	64270709	64283946	20	2526
PYGM-002	ENST00000377432	chr11	-	64270709	64283946	18	2262
PYGM-201	ENST00000436572	chr11	-	64270709	64283946	16	1785

Graphical representation of variants



LEGEND FOR VARIATION TYPES:

spliceOverlap synonymous nonsynonymous prematureStop removedStop insertion deletion substitution

Detailed summary of variants

Chromosome	Position	Reference allele	Alternate allele	Identifier	Type	Fraction of transcripts affected	Transcripts	Details
chr11	64276502	G	C	.	synonymous	3/3	ENST00000164139 ENST00000377432 ENST00000436572	2526_1568_523_L->L 2262_1304_435_L->L 1785_827_276_L->L
chr11	64281910	C	A	.	nonsynonymous	3/3	ENST00000164139 ENST00000377432 ENST00000436572	2526_576_192_A->S 2262_312_104_A->S 1785_312_104_A->S
chr11	64283799	G	A	.	prematureStop	2/3	ENST00000164139 ENST00000377432	2526_147_49_R->* 2262_147_49_R->*

Next steps

- User-specified VCF files (file upload)
- Improvement of web-interface
  - Hyperlinking
- Population statistics for variants