EVA: Exome Variation Analyzer, a convivial tool for filtering strategies

S. Coutant[1,2], A. Lefebvre[2], M. Léonard[2], É. Prieur-Gaston[2], D. Campion[1], T. Lecroq[2] and H. Dauchel[2]

1. University of Rouen, France, INSERM: National Institute of Health and Medical Research
U614: Molecular genetics of cancer and neuropsychiatric diseases
2. University of Rouen, France, LITIS EA 4108: Computer science, information processing and systems laboratory

Identifying relevant genes

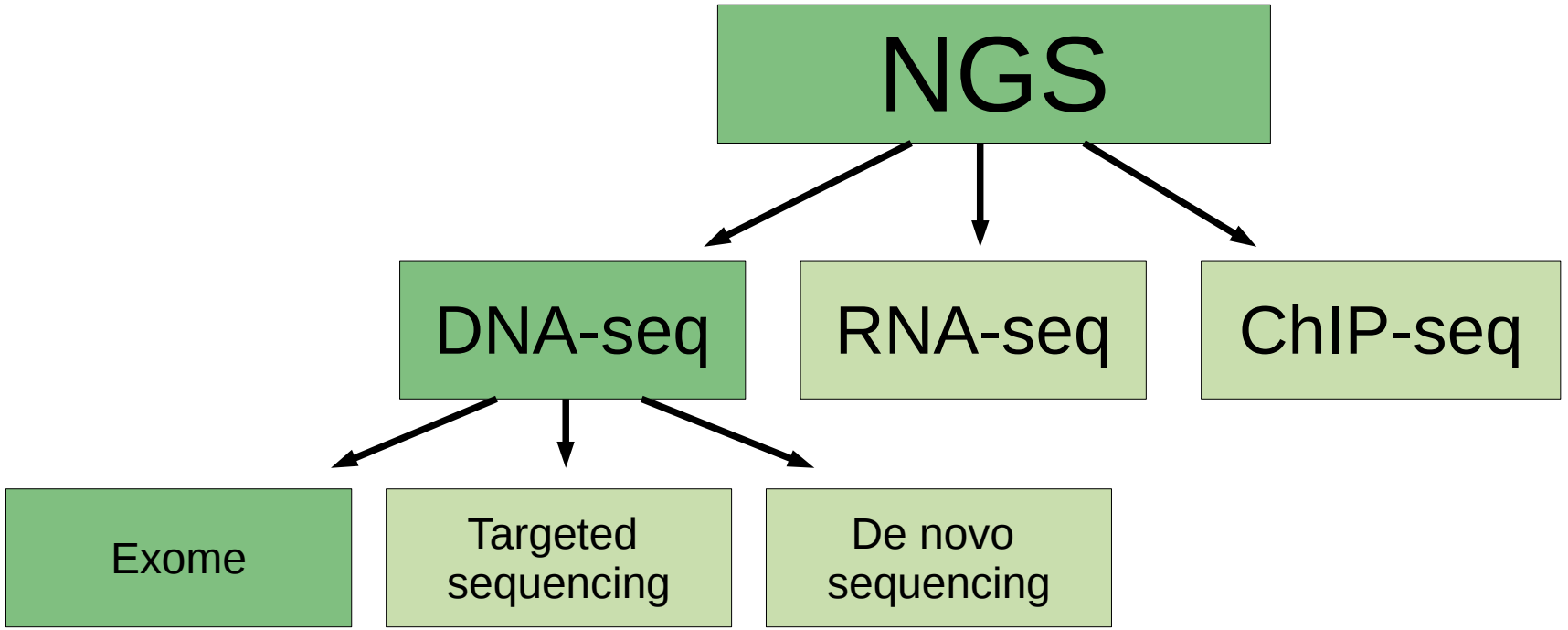
Use of genetic markers :

- Quantitative Trait Locus mapping
- Linkage Analysis
- ...
- Genome-Wide Association Study

→ Molecular basis for nearly 3,000 Mendelian disorders is known

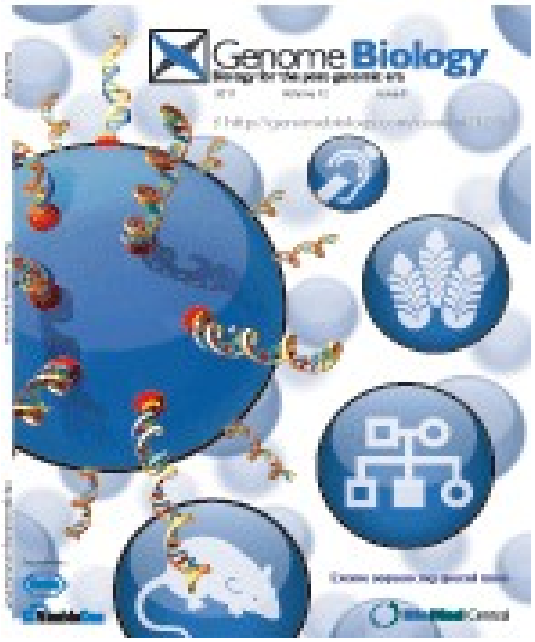
N.O. Stitziel, A. Kiezun & S. Sunyaev. Computational and statistical approaches to analysing variants identified by exome sequencing. *Genome Biology* **12**(9) 2011, 227

NGS: New Generation Sequencing



J. Shendure & H. Ji. Next-generation DNA sequencing. *Nature Biotechnology* **26**(10) (2008) 1135-1145

Exome Sequencing



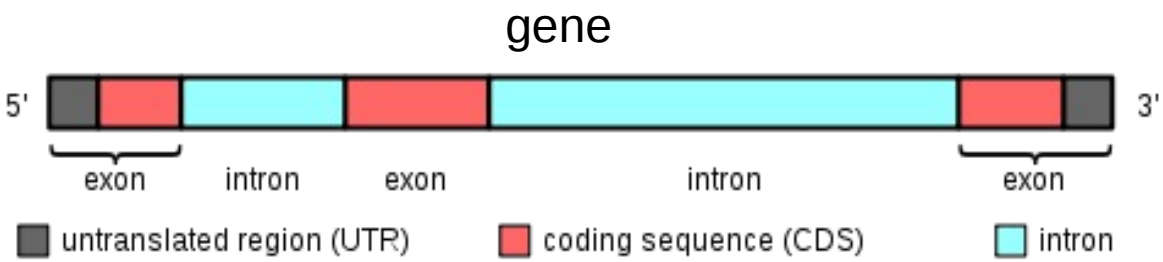
The last issue of *Genome Biology* (volume 12 issue 9, 2011) is completely dedicated to exome sequencing

Exome sequencing in Nature Genetics:
- 2010: 6 studies
- 2011: 18 studies

Editorial. *Nature Genetics* **43** 921 (2011)

Exome

The “exome” represents all the exons in the genome
(ie, the transcribed region of the genes)



Human exome:
· 180,000 exons
· ~30 Mb vs. ~3Gb for the whole genome
· ~1% of the total human genome

Capture

Agilent Technologies

The Agilent SureSelect Human All Exon Kit version 1 captures:
· 180,000 CCDS database (NCBI)
· 700 miRNA
· 300 ncRNA

38Mb (3 µg DNA needed)

Proof of concept

Identifying a gene responsible in a Mendelian disorder was proved possible using whole exome sequencing.

ARTICLES

nature genetics

August 2009

Exome sequencing identifies the cause of a mendelian disorder

Sarah B Ng[1,10], Kati J Buckingham[2,10], Choli Lee[1], Abigail W Bigham[2], Holly K Tabor[2,3], Karin M Dent[4], Chad D Huff[5], Paul T Shannon[6], Ethylin Wang Jabs[7,8], Deborah A Nickerson[1], Jay Shendure[1] & Michael J Bamshad[1,2,9]

Nature. 2009 September 10; 461(7261): 272–276. doi:10.1038/nature08250.

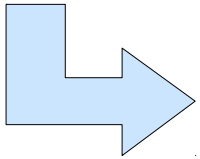
Targeted Capture and Massively Parallel Sequencing of Twelve Human Exomes

Sarah B. Ng[1], Emily H. Turner[1], Peggy D. Robertson[1], Steven D. Flygare[1], Abigail W. Bigham[2], Choli Lee[1], Tristan Shaffer[1], Michelle Wong[1], Arindam Bhattacharjee[3], Evan E. Eichler[1,4], Michael Bamshad[2], Deborah A. Nickerson[1], and Jay Shendure[1]

Recurrence strategy

Exome sequencing:

17,000 cSNPs per individual: 95% in dbSNP
166 indels per individual: 63% in dbSNP



Filters needed

Compare to ~3 million SNPs per individual (in the whole genome)

Recurrence strategy

Exome sequencing:

17,000 cSNPs per individual: 95% in dbSNP
166 indels per individual: 63% in dbSNP

Filters needed

Number of genes affected by at least one cSNP in 1 2 3 4 individuals

	1	2	3	4
Nonsynonymous cSNP	4,510	3,284	2,765	2,479
Not in dbSNP	513	128	71	53
Not in HapMap	799	168	53	21
Not in dbSNP + HapMap	360	38	8	1 (MYH3)
Predicted damaging	160	10	2	1 (MYH3)

Fig2 : From Ng S B, *et al.* Nature 461, 272-276 (2009). [1]

Recurrence strategy

Exome sequencing :

17,000 cSNPs by individual: 95% in dbSNP
166 indels by individual: 63% in dbSNP

Filters needed

Number of genes affected by at least one cSNP in	1	2	3	4 individuals
Nonsynonymous cSNP	4,510	3,284	2,765	2,479
Not in dbSNP	513	128	71	53
Not in HapMap	799	168	53	21
Not in dbSNP + HapMap	360	38	8	1 (MYH3)
Predicted damaging	160	10	2	1 (MYH3)

Freeman-Sheldon syndrome

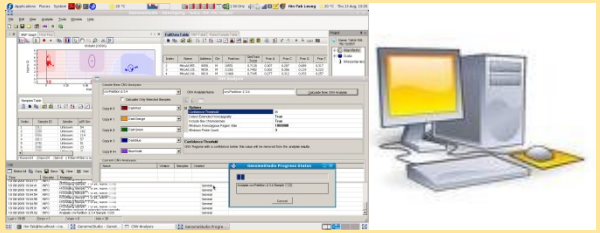
Fig2 : From Ng S B, *et al.* Nature 461, 272-276 (2009). [1]

Problematic: clinical bioinformatics

NGS sequencing

Illumina - GA IIx

Mapping & variations detection



CASAVA + bioinformatics processing

?

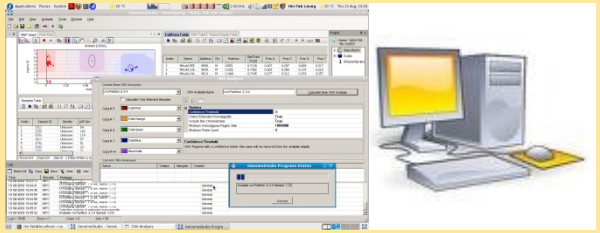
INTEGRAGEN

Problematic

NGS sequencing

Illumina - GA IIx

Mapping & variations detection



CASAVA + bioinformatics processing

We need to
Filter variations
To make the clinician
Autonomous
And to make a step towards
Personalized medecine

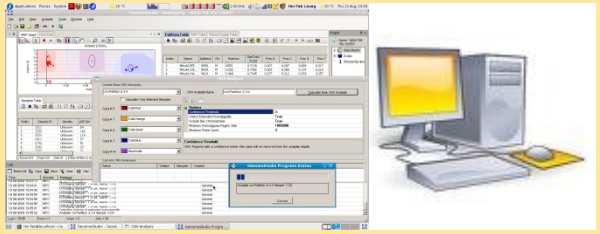
INTEGRAGEN

EVA - Exome Variation Analyzer

NGS sequencing

Illumina - GA IIx

Mapping & variations detection



CASAVA + bioinformatics processing

EVA integration module

ExomeDB

EVA

INTEGRAGEN

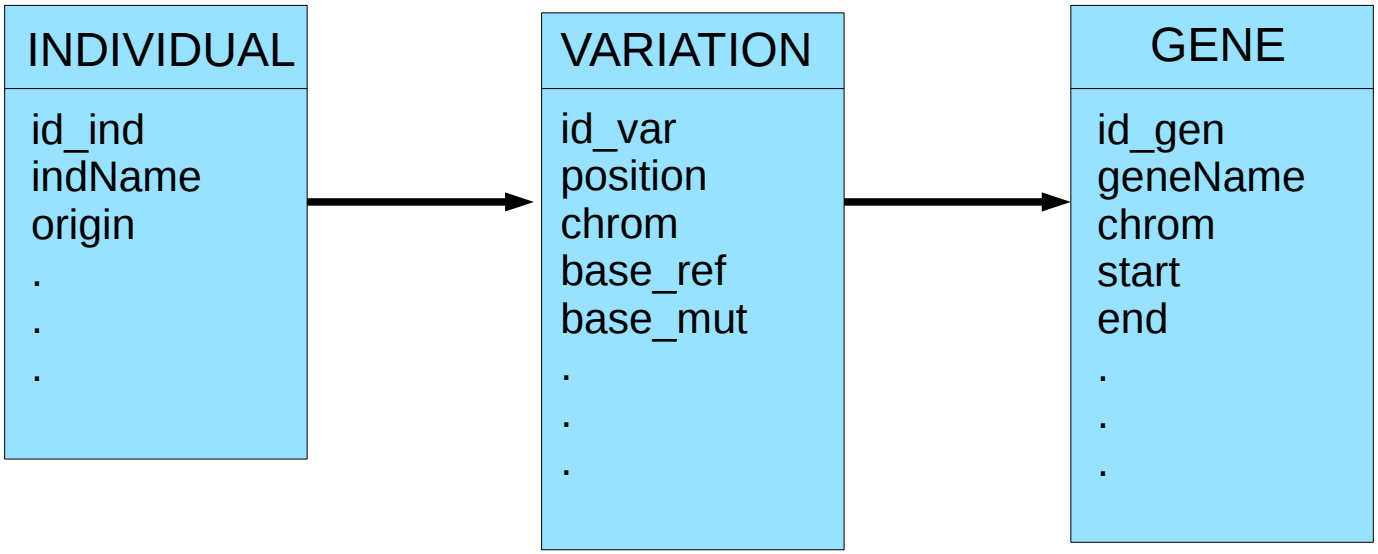
The EVA tool consists of:

- a database: ExomeDB
- a browser
- several filters and search tools

Database: ExomeDB

Structure

- Developed in mySQL (ver 5.0)
- Principal tables: Individual, Variation and Gene



Integration module

- Every new project is subject to a remote loading using an online integration module. This module accepts .txt files and .xls files

- The integrated data are: lists of variations (SNP, InDel) + their annotations (position, mutation type, ...)

- Output of a CASAVA-like analysis pipeline. The tool is optimised to admit data coming from IntegraGen, biotechnology society, Évry, France

Position.Build36	sample.ID	chrom	A	C	G	T	modified_call	total	used	score	reference	type
59374	ALZ 9002	chr1	0	0	49	0	G	74	49	161.88	A	SNP_diff
866362	ALZ 9002	chr1	0	0	2	0	G	2	2	6.40	A	SNP_diff.douteux
867694	ALZ 9002	chr1	0	3	0	0	C	3	3	7.79	T	SNP_diff.douteux
870253	ALZ 9002	chr1	17	9	0	0	AC	26	26	56.50:28.20	C	SNP_het2
873488	ALZ 9002	chr1	0	0	3	0	G	5	3	8.80	A	SNP_diff.douteux
877423	ALZ 9002	chr1	0	13	0	0	C	19	13	43.80	A	SNP_diff
877664	ALZ 9002	chr1	0	0	11	0	G	11	11	39.70	A	SNP_diff
878502	ALZ 9002	chr1	0	27	0	0	C	32	27	91.49	T	SNP_diff
878522	ALZ 9002	chr1	0	17	0	0	C	22	17	54.78	T	SNP_diff

rs.name	Gene.name	Gene.start	Gene.end	strand	nbre.exon	refseq	typeannot	type.pos
rs2691305	OR4F5	58953	59871	0	1	NM_001005484	in-cds-fauxsens	Exon 1
rs4372192	SAMD11	850983	869824	0	14	NM_152486	in-intron	Intron 7
rs6672356	SAMD11	850983	869824	0	14	NM_152486	in-cds-fauxsens	Exon 10
rs3748593	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 18
rs4970378	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 13
rs3748595	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 10
rs3828047	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 10
rs3748596	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 9
rs3748597	NOC2L	869445	884542	-0	19	NM_015658	in-cds-fauxsens	Exon 9

Integration module

- Every new project is subject to a remote loading using an online integration module. This module accepts .txt files and .xls files

- The integrated data are: lists of variations (SNP, InDel) + their annotations (position, mutation type, ...)

- Output of a CASAVA-like analysis pipeline. The tool is optimised to admit data coming from IntegraGen, biotechnology society, Évry, France

Genomic position

Position.Build36	sample.ID	chrom	A	C	G	T	modified_call	total	used	score	reference	type
59374	ALZ 9002	chr1	0	0	49	0	G	74	49	161.88	A	SNP_diff
866362	ALZ 9002	chr1	0	0	2	0	G	2	2	6.40	A	SNP_diff.douteux
867694	ALZ 9002	chr1	0	3	0	0	C	3	3	7.79	T	SNP_diff.douteux
870253	ALZ 9002	chr1	17	9	0	0	AC	26	26	56.50:28.20	C	SNP_het2
873488	ALZ 9002	chr1	0	0	3	0	G	5	3	8.80	A	SNP_diff.douteux
877423	ALZ 9002	chr1	0	13	0	0	C	19	13	43.80	A	SNP_diff
877664	ALZ 9002	chr1	0	0	11	0	G	11	11	39.70	A	SNP_diff
878502	ALZ 9002	chr1	0	27	0	0	C	32	27	91.49	T	SNP_diff
878522	ALZ 9002	chr1	0	17	0	0	C	22	17	54.78	T	SNP_diff

rs.name	Gene.name	Gene.start	Gene.end	strand	nbre.exon	refseq	typeannot	type.pos
rs2691305	OR4F5	58953	59871	0	1	NM_001005484	in-cds-fauxsens	Exon 1
rs4372192	SAMD11	850983	869824	0	14	NM_152486	in-intron	Intron 7
rs6672356	SAMD11	850983	869824	0	14	NM_152486	in-cds-fauxsens	Exon 10
rs3748593	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 18
rs4970378	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 13
rs3748595	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 10
rs3828047	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 10
rs3748596	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 9
rs3748597	NOC2L	869445	884542	-0	19	NM_015658	in-cds-fauxsens	Exon 9

Integration module

- Every new project is subject to a remote loading using an online integration module. This module accepts .txt files and .xls files

- The integrated data are: lists of variations (SNP, InDel) + their annotations (position, mutation type, ...)

- Output of a CASAVA-like analysis pipeline. The tool is optimised to admit data coming from IntegraGen, biotechnology society, Évry, France

Number of read bases

Position.Build36	sample.ID	chrom	A	C	G	T	modified_call	total	used	score	reference	type
59374	ALZ 9002	chr1	0	0	49	0	G	74	49	161.88	A	SNP_diff
866362	ALZ 9002	chr1	0	0	2	0	G	2	2	6.40	A	SNP_diff.douteux
867694	ALZ 9002	chr1	0	3	0	0	C	3	3	7.79	T	SNP_diff.douteux
870253	ALZ 9002	chr1	17	9	0	0	AC	26	26	56.50:28.20	C	SNP_het2
873488	ALZ 9002	chr1	0	0	3	0	G	5	3	8.80	A	SNP_diff.douteux
877423	ALZ 9002	chr1	0	13	0	0	C	19	13	43.80	A	SNP_diff
877664	ALZ 9002	chr1	0	0	11	0	G	11	11	39.70	A	SNP_diff
878502	ALZ 9002	chr1	0	27	0	0	C	32	27	91.49	T	SNP_diff
878522	ALZ 9002	chr1	0	17	0	0	C	22	17	54.78	T	SNP_diff

CONFIDENTIAL

rs.name	Gene.name	Gene.start	Gene.end	strand	nbre.exon	refseq	typeannot	type.pos
rs2691305	OR4F5	58953	59871	0	1	NM_001005484	in-cds-fauxsens	Exon 1
rs4372192	SAMD11	850983	869824	0	14	NM_152486	in-intron	Intron 7
rs6672356	SAMD11	850983	869824	0	14	NM_152486	in-cds-fauxsens	Exon 10
rs3748593	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 18
rs4970378	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 13
rs3748595	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 10
rs3828047	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 10
rs3748596	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 9
rs3748597	NOC2L	869445	884542	-0	19	NM_015658	in-cds-fauxsens	Exon 9

Integration module

- Every new project is subject to a remote loading using an online integration module. This module accepts .txt files and .xls files

- The integrated data are: lists of variations (SNP, InDel) + their annotations (position, mutation type, ...)

- Output of a CASAVA-like analysis pipeline. The tool is optimised to admit data coming from IntegraGen, biotechnology society, Évry, France

Quality and coverage

Position.Build36	sample.ID	chrom	A	C	G	T	modified_call	total	used	score	reference	type
59374	ALZ 9002	chr1	0	0	49	0	G	74	49	161.88	A	SNP_diff
866362	ALZ 9002	chr1	0	0	2	0	G	2	2	6.40	A	SNP_diff.douteux
867694	ALZ 9002	chr1	0	3	0	0	C	3	3	7.79	T	SNP_diff.douteux
870253	ALZ 9002	chr1	17	9	0	0	AC	26	26	56.50:28.20	C	SNP_het2
873488	ALZ 9002	chr1	0	0	3	0	G	5	3	8.80	A	SNP_diff.douteux
877423	ALZ 9002	chr1	0	13	0	0	C	19	13	43.80	A	SNP_diff
877664	ALZ 9002	chr1	0	0	11	0	G	11	11	39.70	A	SNP_diff
878502	ALZ 9002	chr1	0	27	0	0	C	32	27	91.49	T	SNP_diff
878522	ALZ 9002	chr1	0	17	0	0	C	22	17	54.78	T	SNP_diff

CONFIDENTIAL

rs.name	Gene.name	Gene.start	Gene.end	strand	nbre.exon	refseq	typeannot	type.pos
rs2691305	OR4F5	58953	59871	0	1	NM_001005484	in-cds-fauxsens	Exon 1
rs4372192	SAMD11	850983	869824	0	14	NM_152486	in-intron	Intron 7
rs6672356	SAMD11	850983	869824	0	14	NM_152486	in-cds-fauxsens	Exon 10
rs3748593	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 18
rs4970378	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 13
rs3748595	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 10
rs3828047	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 10
rs3748596	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 9
rs3748597	NOC2L	869445	884542	-0	19	NM_015658	in-cds-fauxsens	Exon 9

Integration module

• Every new project is subject to a remote loading using an online integration module. This module accepts .txt files and .xls files

• The integrated data are: lists of variations (SNP, InDel) + their annotations (position, mutation type, ...)

• Output of a CASAVA-like analysis pipeline. The tool is optimised to admit data coming from IntegraGen, biotechnology society, Évry, France

Mutated base / reference base

Position.Build36	sample.ID	chrom	A	C	G	T	modified_call	total	used	score	reference	type
59374	ALZ 9002	chr1	0	0	49	0	G	74	49	161.88	A	SNP_diff
866362	ALZ 9002	chr1	0	0	2	0	G	2	2	6.40	A	SNP_diff.douteux
867694	ALZ 9002	chr1	0	3	0	0	C	3	3	7.79	T	SNP_diff.douteux
870253	ALZ 9002	chr1	17	9	0	0	AC	26	26	56.50:28.20	C	SNP_het2
873488	ALZ 9002	chr1	0	0	3	0	G	5	3	8.80	A	SNP_diff.douteux
877423	ALZ 9002	chr1	0	13	0	0	C	19	13	43.80	A	SNP_diff
877664	ALZ 9002	chr1	0	0	11	0	G	11	11	39.70	A	SNP_diff
878502	ALZ 9002	chr1	0	27	0	0	C	32	27	91.49	T	SNP_diff
878522	ALZ 9002	chr1	0	17	0	0	C	22	17	54.78	T	SNP_diff

CONFIDENTIAL

rs.name	Gene.name	Gene.start	Gene.end	strand	nbre.exon	refseq	typeannot	type.pos
rs2691305	OR4F5	58953	59871	0	1	NM_001005484	in-cds-fauxsens	Exon 1
rs4372192	SAMD11	850983	869824	0	14	NM_152486	in-intron	Intron 7
rs6672356	SAMD11	850983	869824	0	14	NM_152486	in-cds-fauxsens	Exon 10
rs3748593	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 18
rs4970378	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 13
rs3748595	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 10
rs3828047	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 10
rs3748596	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 9
rs3748597	NOC2L	869445	884542	-0	19	NM_015658	in-cds-fauxsens	Exon 9

Integration module

- Every new project is subject to a remote loading using an online integration module. This module accepts .txt files and .xls files

- The integrated data are: lists of variations (SNP, InDel) + their annotations (position, mutation type, ...)

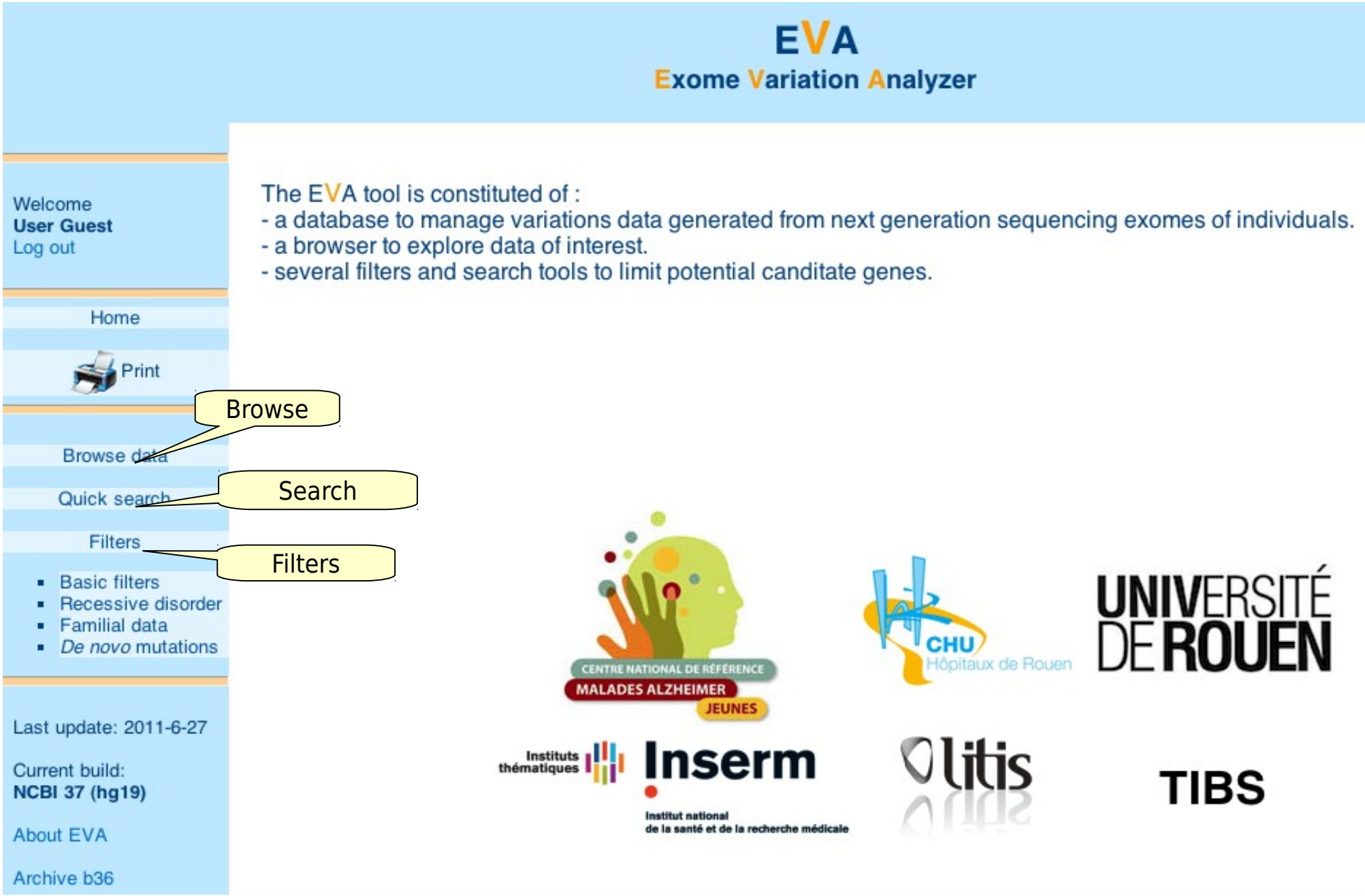
- Output of a CASAVA-like analysis pipeline. The tool is optimised to admit data coming from IntegraGen, biotechnology society, Évry, France

Gene annotations: gene name and functional class

Position.Build36	sample.ID	chrom	A	C	G	T	modified_call	total	used	score	reference	type
59374	ALZ 9002	chr1	0	0	49	0	G	74	49	161.88	A	SNP_diff
866362	ALZ 9002	chr1	0	0	2	0	G	2	2	6.40	A	SNP_diff.douteux
867694	ALZ 9002	chr1	0	3	0	0	C	3	3	7.79	T	SNP_diff.douteux
870253	ALZ 9002	chr1	17	9	0	0	AC	26	26	56.50:28.20	C	SNP_het2
873488	ALZ 9002	chr1	0	0	3	0	G	5	3	8.80	A	SNP_diff.douteux
877423	ALZ 9002	chr1	0	13	0	0	C	19	13	43.80	A	SNP_diff
877664	ALZ 9002	chr1	0	0	11	0	G	11	11	39.70	A	SNP_diff
878502	ALZ 9002	chr1	0	27	0	0	C	32	27	91.49	T	SNP_diff
878522	ALZ 9002	chr1	0	17	0	0	C	22	17	54.78	T	SNP_diff

rs.name	Gene.name	Gene.start	Gene.end	strand	nbre.exon	refseq	typeannot	type.pos
rs2691305	OR4F5	58953	59871	0	1	NM_001005484	in-cds-fauxsens	Exon 1
rs4372192	SAMD11	850983	869824	0	14	NM_152486	in-intron	Intron 7
rs6672356	SAMD11	850983	869824	0	14	NM_152486	in-cds-fauxsens	Exon 10
rs3748593	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 18
rs4970378	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 13
rs3748595	NOC2L	869445	884542	-0	19	NM_015658	in-intron	Intron 10
rs3828047	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 10
rs3748596	NOC2L	869445	884542	-0	19	NM_015658	in-cds-silencieuse	Exon 9
rs3748597	NOC2L	869445	884542	-0	19	NM_015658	in-cds-fauxsens	Exon 9

Web Interface



14 exomes in early autosomic dominant Alzheimer pathology without identified mutations

[Variations overview]

Individual	Known variations								Unknown variations								Total
	Exon						Intron	Sub-total	Exon						Intron	Sub-total	
	Single variation				Indel		Splice		Single variation				Indel		Splice		
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		
ALZ 049	7739	6301	9	30	21	19	78	**14197**	347	567	0	14	60	57	9	**1054**	**15251**
ALZ 426	8030	6534	7	30	20	19	76	**14716**	333	526	1	12	62	54	6	**994**	**15710**
ROU 632	8040	6540	3	34	19	18	82	**14736**	323	527	2	20	54	71	19	**1016**	**15752**
EXT 049	8060	6696	5	29	18	22	68	**14898**	382	602	1	14	71	74	14	**1158**	**16056**
EXT 055	7747	6210	7	32	19	21	68	**14104**	359	623	0	23	65	59	13	**1142**	**15246**
ALZ 062	7876	6385	7	33	22	18	74	**14415**	345	594	1	13	71	58	6	**1088**	**15503**
ROU 816	8011	6527	5	39	15	23	81	**14701**	362	587	1	13	73	57	11	**1104**	**15805**
ALZ 198	7282	5930	5	27	19	22	66	**13351**	314	575	1	12	56	71	10	**1039**	**14390**
ALZ 056	7860	6592	5	33	22	14	74	**14600**	280	522	2	15	40	57	18	**934**	**15534**
EXT 094	8300	6837	5	41	19	19	91	**15312**	338	563	2	10	49	56	10	**1028**	**16340**
EXT 077	8050	6641	7	39	23	17	74	**14851**	309	478	2	9	53	51	9	**911**	**15762**
EXT 050	8156	6683	7	27	21	20	75	**14989**	274	459	0	10	53	58	15	**869**	**15858**
EXT 220	10070	8558	26	36	21	19	94	**18824**	362	585	1	2	152	115	14	**1231**	**20055**
EXT 181	9981	8487	27	30	19	16	84	**18644**	373	681	3	4	167	107	22	**1357**	**20001**

14 exomes in early autosomic dominant Alzheimer pathology without identified mutations

[Variations overview]

Individual	Known variations								Unknown variations								Total
	Exon						Intron	Sub-total	Exon						Intron	Sub-total	
	Single variation				Indel		Splice		Single variation				Indel		Splice		
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		
ALZ 049	7739	6301	9	30	21	19	78	**14197**	347	567	0	14	60	57	9	**1054**	**15251**
ALZ 426	8030	6534	7	30	20	19	76	**14716**	333	526	1	12	62	54	6	**994**	**15710**
ROU 632	8040	6540	3	34	19	18	82	**14736**	323	527	2	20	54	71	19	**1016**	**15752**
EXT 049	8060	6696	5	29	18	22	68	**14898**	382	602	1	14	71	74	14	**1158**	**16056**
EXT 055	7747	6210	7	32	19	21	68	**14104**	359	623	0	23	65	59	13	**1142**	**15246**
ALZ 062	7876	6385	7	33	22	18	74	**14415**	345	594	1	13	71	58	6	**1088**	**15503**
ROU 816	8011	6527	5	39	15	23	81	**14701**	362	587	1	13	73	57	11	**1104**	**15805**
ALZ 198	7282	5930	5	27	19	22	66	**13351**	314	575	1	12	56	71	10	**1039**	**14390**
ALZ 056	7860	6592	5	33	22	14	74	**14600**	280	522	2	15	40	57	18	**934**	**15534**
EXT 094	8300	6837	5	41	19	19	91	**15312**	338	563	2	10	49	56	10	**1028**	**16340**
EXT 077	8050	6641	7	39	23	17	74	**14851**	309	478	2	9	53	51	9	**911**	**15762**
EXT 050	8156	6683	7	27	21	20	75	**14989**	274	459	0	10	53	58	15	**869**	**15858**
EXT 220	10070	8558	26	36	21	19	94	**18824**	362	585	1	2	152	115	14	**1231**	**20055**
EXT 181	9981	8487	27	30	19	16	84	**18644**	373	681	3	4	167	107	22	**1357**	**20001**

Sequenced individuals

14 exomes in early autosomic dominant Alzheimer pathology without identified mutations

[Variations overview]

Individual	Known variations								Unknown variations								Total
	Exon						Intron	Sub-total	Exon						Intron	Sub-total	
	Single variation				Indel		Splice		Single variation				Indel		Splice		
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		
ALZ 049	7739	6301	9	30	21	19	78	**14197**	347	567	0	14	60	57	9	**1054**	**15251**
ALZ 426	8030	6534	7	30	20	19	76	**14716**	333	526	1	12	62	54	6	**994**	**15710**
ROU 632	8040	6540	3	34	19	18	82	**14736**	323	527	2	20	54	71	19	**1016**	**15752**
EXT 049	8060	6696	5	29	18	22	68	**14898**	382	602	1	14	71	74	14	**1158**	**16056**
EXT 055	7747	6210	7	32	19	21	68	**14104**	359	623	0	23	65	59	13	**1142**	**15246**
ALZ 062	7876	6385	7	33	22	18	74	**14415**	345	594	1	13	71	58	6	**1088**	**15503**
ROU 816	8011	6527	5	39	15	23	81	**14701**	362	587	1	13	73	57	11	**1104**	**15805**
ALZ 198	7282	5930	5	27	19	22	66	**13351**	314	575	1	12	56	71	10	**1039**	**14390**
ALZ 056	7860	6592	5	33	22	14	74	**14600**	280	522	2	15	40	57	18	**934**	**15534**
EXT 094	8300	6837	5	41	19	19	91	**15312**	338	563	2	10	49	56	10	**1028**	**16340**
EXT 077	8050	6641	7	39	23	17	74	**14851**	309	478	2	9	53	51	9	**911**	**15762**
EXT 050	8156	6683	7	27	21	20	75	**14989**	274	459	0	10	53	58	15	**869**	**15858**
EXT 220	10070	8558	26	36	21	19	94	**18824**	362	585	1	2	152	115	14	**1231**	**20055**
EXT 181	9981	8487	27	30	19	16	84	**18644**	373	681	3	4	167	107	22	**1357**	**20001**

In dbSNP

Not in dbSNP

14 exomes in early autosomic dominant Alzheimer pathology without identified mutations

[Variations overview]

Individual	Known variations								Unknown variations								Total
	Exon						Intron	Sub-total	Exon						Intron	Sub-total	
	Single variation				Indel		Splice		Single variation				Indel		Splice		
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		
ALZ 049	7739	6301	9	30	21	19	78	**14197**	347	567	0	14	60	57	9	**1054**	**15251**
ALZ 426	8030	6534	7	30	20	19	76	**14716**	333	526	1	12	62	54	6	**994**	**15710**
ROU 632	8040	6540	3	34	19	18	82	**14736**	323	527	2	20	54	71	19	**1016**	**15752**
EXT 049	8060	6696	5	29	18	22	68	**14898**	382	602	1	14	71	74	14	**1158**	**16056**
EXT 055	7747	6210	7	32	19	21	68	**14104**	359	623	0	23	65	59	13	**1142**	**15246**
ALZ 062	7876	6385	7	33	22	18	74	**14415**	345	594	1	13	71	58	6	**1088**	**15503**
ROU 816	8011	6527	5	39	15	23	81	**14701**	362	587	1	13	73	57	11	**1104**	**15805**
ALZ 198	7282	5930	5	27	19	22	66	**13351**	314	575	1	12	56	71	10	**1039**	**14390**
ALZ 056	7860	6592	5	33	22	14	74	**14600**	280	522	2	15	40	57	18	**934**	**15534**
EXT 094	8300	6837	5	41	19	19	91	**15312**	338	563	2	10	49	56	10	**1028**	**16340**
EXT 077	8050	6641	7	39	23	17	74	**14851**	309	478	2	9	53	51	9	**911**	**15762**
EXT 050	8156	6683	7	27	21	20	75	**14989**	274	459	0	10	53	58	15	**869**	**15858**
EXT 220	10070	8558	26	36	21	19	94	**18824**	362	585	1	2	152	115	14	**1231**	**20055**
EXT 181	9981	8487	27	30	19	16	84	**18644**	373	681	3	4	167	107	22	**1357**	**20001**

Exonic / Intronic

14 exomes in early autosomic dominant Alzheimer pathology without identified mutations

[Variations overview]

Individual	Known variations								Unknown variations								Total
	Exon						Intron	Sub-total	Exon						Intron	Sub-total	
	Single variation				Indel		Splice		Single variation				Indel		Splice		
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		
ALZ 049	7739	6301	9	30	21	19	78	**14197**	347	567	0	14	60	57	9	**1054**	**15251**
ALZ 426	8030	6534	7	30	20	19	76	**14716**	333	526	1	12	62	54	6	**994**	**15710**
ROU 632	8040	6540	3	34	19	18	82	**14736**	323	527	2	20	54	71	19	**1016**	**15752**
EXT 049	8060	6696	5	29	18	22	68	**14898**	382	602	1	14	71	74	14	**1158**	**16056**
EXT 055	7747	6210	7	32	19	21	68	**14104**	359	623	0	23	65	59	13	**1142**	**15246**
ALZ 062	7876	6385	7	33	22	18	74	**14415**	345	594	1	13	71	58	6	**1088**	**15503**
ROU 816	8011	6527	5	39	15	23	81	**14701**	362	587	1	13	73	57	11	**1104**	**15805**
ALZ 198	7282	5930	5	27	19	22	66	**13351**	314	575	1	12	56	71	10	**1039**	**14390**
ALZ 056	7860	6592	5	33	22	14	74	**14600**	280	522	2	15	40	57	18	**934**	**15534**
EXT 094	8300	6837	5	41	19	19	91	**15312**	338	563	2	10	49	56	10	**1028**	**16340**
EXT 077	8050	6641	7	39	23	17	74	**14851**	309	478	2	9	53	51	9	**911**	**15762**
EXT 050	8156	6683	7	27	21	20	75	**14989**	274	459	0	10	53	58	15	**869**	**15858**
EXT 220	10070	8558	26	36	21	19	94	**18824**	362	585	1	2	152	115	14	**1231**	**20055**
EXT 181	9981	8487	27	30	19	16	84	**18644**	373	681	3	4	167	107	22	**1357**	**20001**

Single variation / Insertion - deletion

14 exomes in early autosomic dominant Alzheimer pathology without identified mutations

[Variations overview]

Individual	Known variations								Unknown variations								Total
	Exon				Indel		Intron Splice	Sub-total	Exon				Indel		Intron Splice	Sub-total	
	Single variation								Single variation								
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		
ALZ 049	7739	6301	9	30	21	19	78	**14197**	347	567	0	14	60	57	9	**1054**	**15251**
ALZ 426	8030	6534	7	30	20	19	76	**14716**	333	526	1	12	62	54	6	**994**	**15710**
ROU 632	8040	6540	3	34	19	18	82	**14736**	323	527	2	20	54	71	19	**1016**	**15752**
EXT 049	8060	6696	5	29	18	22	68	**14898**	382	602	1	14	71	74	14	**1158**	**16056**
EXT 055	7747	6210	7	32	19	21	68	**14104**	359	623	0	23	65	59	13	**1142**	**15246**
ALZ 062	7876	6385	7	33	22	18	74	**14415**	345	594	1	13	71	58	6	**1088**	**15503**
ROU 816	8011	6527	5	39	15	23	81	**14701**	362	587	1	13	73	57	11	**1104**	**15805**
ALZ 198	7282	5930	5	27	19	22	66	**13351**	314	575	1	12	56	71	10	**1039**	**14390**
ALZ 056	7860	6592	5	33	22	14	74	**14600**	280	522	2	15	40	57	18	**934**	**15534**
EXT 094	8300	6837	5	41	19	19	91	**15312**	338	563	2	10	49	56	10	**1028**	**16340**
EXT 077	8050	6641	7	39	23	17	74	**14851**	309	478	2	9	53	51	9	**911**	**15762**
EXT 050	8156	6683	7	27	21	20	75	**14989**	274	459	0	10	53	58	15	**869**	**15858**
EXT 220	10070	8558	26	36	21	19	94	**18824**	362	585	1	2	152	115	14	**1231**	**20055**
EXT 181	9981	8487	27	30	19	16	84	**18644**	373	681	3	4	167	107	22	**1357**	**20001**

Single variation categories:
Synonym - Missense - Stop loss - Nonsense

Filters
Recurrence Strategy - 1st step: select project

14 exomes in early autosomic dominant Alzheimer pathology without identified mutations

[Variations overview]

Individual	Known variations								Unknown variations								Total
	Exon						Intron	Sub-total	Exon						Intron	Sub-total	
	Single variation				Indel		Splice		Single variation				Indel		Splice		
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		
ALZ 049	7739	6301	9	30	21	19	78	**14197**	347	567	0	14	60	57	9	**1054**	**15251**
ALZ 426	8030	6534	7	30	20	19	76	**14716**	333	526	1	12	62	54	6	**994**	**15710**
ROU 632	8040	6540	3	34	19	18	82	**14736**	323	527	2	20	54	71	19	**1016**	**15752**
EXT 049	8060	6696	5	29	18	22	68	**14898**	382	602	1	14	71	74	14	**1158**	**16056**
EXT 055	7747	6210	7	32	19	21	68	**14104**	359	623	0	23	65	59	13	**1142**	**15246**
ALZ 062	7876	6385	7	33	22	18	74	**14415**	345	594	1	13	71	58	6	**1088**	**15503**
ROU 816	8011	6527	5	39	15	23	81	**14701**	362	587	1	13	73	57	11	**1104**	**15805**
ALZ 198	7282	5930	5	27	19	22	66	**13351**	314	575	1	12	56	71	10	**1039**	**14390**
ALZ 056	7860	6592	5	33	22	14	74	**14600**	280	522	2	15	40	57	18	**934**	**15534**
EXT 094	8300	6837	5	41	19	19	91	**15312**	338	563	2	10	49	56	10	**1028**	**16340**
EXT 077	8050	6641	7	39	23	17	74	**14851**	309	478	2	9	53	51	9	**911**	**15762**
EXT 050	8156	6683	7	27	21	20	75	**14989**	274	459	0	10	53	58	15	**869**	**15858**
EXT 220	10070	8558	26	36	21	19	94	**18824**	362	585	1	2	152	115	14	**1231**	**20055**
EXT 181	9981	8487	27	30	19	16	84	**18644**	373	681	3	4	167	107	22	**1357**	**20001**

Indel categories:
Frameshift - No Frameshift

14 exomes in early autosomic dominant Alzheimer pathology without identified mutations

[Variations overview]

Individual	Known variations								Unknown variations								Total
	Exon						Intron	Sub-total	Exon						Intron	Sub-total	
	Single variation				Indel		Splice		Single variation				Indel		Splice		
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		
ALZ 049	7739	6301	9	30	21	19	78	**14197**	347	567	0	14	60	57	9	**1054**	**15251**
ALZ 426	8030	6534	7	30	20	19	76	**14716**	333	526	1	12	62	54	6	**994**	**15710**
ROU 632	8040	6540	3	34	19	18	82	**14736**	323	527	2	20	54	71	19	**1016**	**15752**
EXT 049	8060	6696	5	29	18	22	68	**14898**	382	602	1	14	71	74	14	**1158**	**16056**
EXT 055	7747	6210	7	32	19	21	68	**14104**	359	623	0	23	65	59	13	**1142**	**15246**
ALZ 062	7876	6385	7	33	22	18	74	**14415**	345	594	1	13	71	58	6	**1088**	**15503**
ROU 816	8011	6527	5	39	15	23	81	**14701**	362	587	1	13	73	57	11	**1104**	**15805**
ALZ 198	7282	5930	5	27	19	22	66	**13351**	314	575	1	12	56	71	10	**1039**	**14390**
ALZ 056	7860	6592	5	33	22	14	74	**14600**	280	522	2	15	40	57	18	**934**	**15534**
EXT 094	8300	6837	5	41	19	19	91	**15312**	338	563	2	10	49	56	10	**1028**	**16340**
EXT 077	8050	6641	7	39	23	17	74	**14851**	309	478	2	9	53	51	9	**911**	**15762**
EXT 050	8156	6683	7	27	21	20	75	**14989**	274	459	0	10	53	58	15	**869**	**15858**
EXT 220	10070	8558	26	36	21	19	94	**18824**	362	585	1	2	152	115	14	**1231**	**20055**
EXT 181	9981	8487	27	30	19	16	84	**18644**	373	681	3	4	167	107	22	**1357**	**20001**

Canonical splice site mutation

14 exomes in early autosomic dominant Alzheimer pathology without identified mutations

[Variations overview]

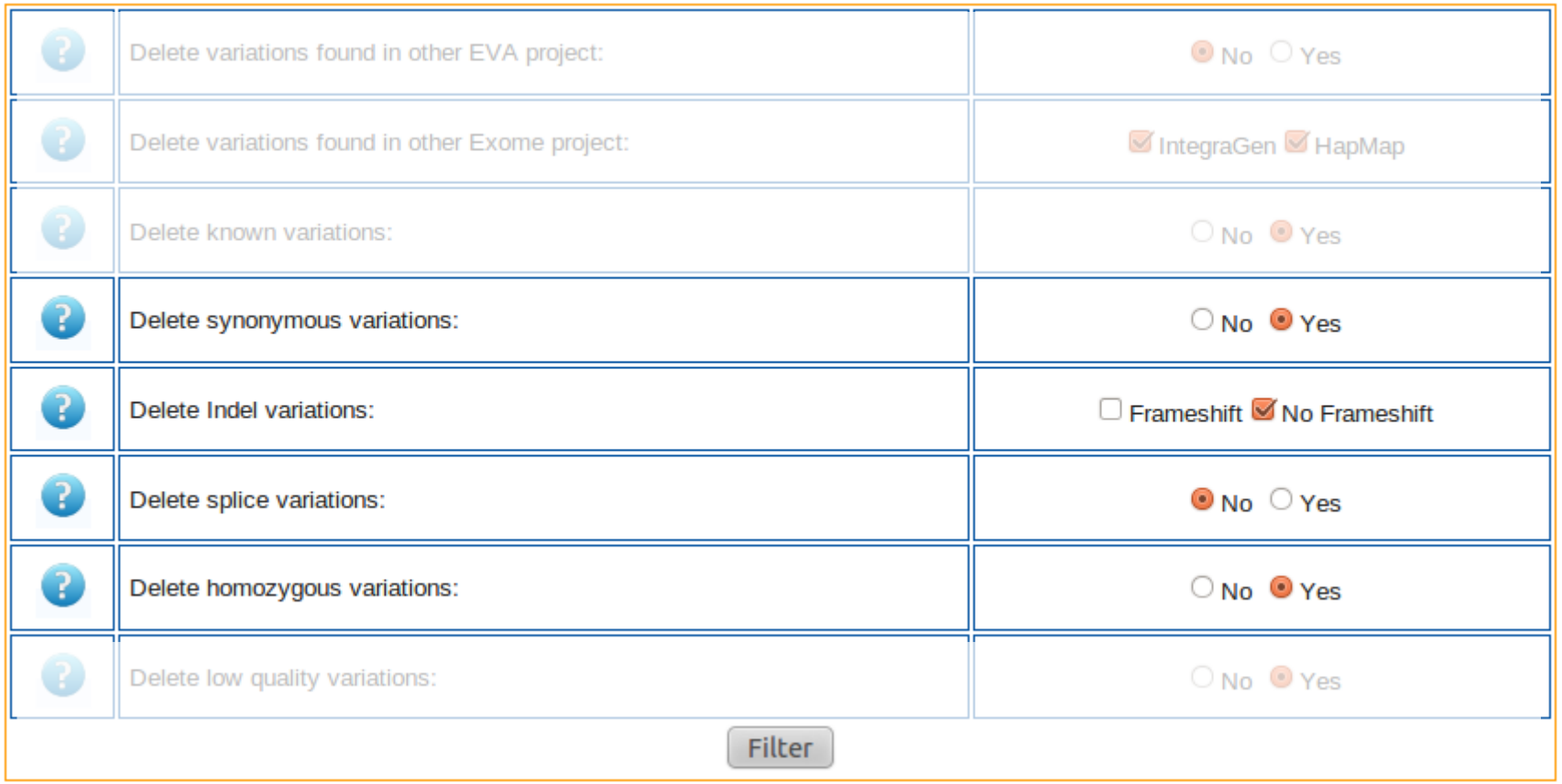
Individual	Known variations								Unknown variations								Total
	Exon						Intron	Sub-total	Exon						Intron	Sub-total	
	Single variation				Indel		Splice		Single variation				Indel		Splice		
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		
ALZ 049	7739	6301	9	30	21	19	78	14197	347	567	0	14	60	57	9	1054	15251
ALZ 426	8030	6534	7	30	20	19	76	14716	333	526	1	12	62	54	6	994	15710
ROU 632	8040	6540	3	34	19	18	82	14736	323	527	2	20	54	71	19	1016	15752
EXT 049	8060	6696	5	29	18	22	68	14898	382	602	1	14	71	74	14	1158	16056
EXT 055	7747	6210	7	32	19	21	68	14104	359	623	0	23	65	59	13	1142	15246
ALZ 062	7876	6385	7	33	22	18	74	14415	345	594	1	13	71	58	6	1088	15503
ROU 816	8011	6527	5	39	15	23	81	14701	362	587	1	13	73	57	11	1104	15805
ALZ 198	7282	5930	5	27	19	22	66	13351	314	575	1	12	56	71	10	1039	14390
ALZ 056	7860	6592	5	33	22	14	74	14600	280	522	2	15	40	57	18	934	15534
EXT 094	8300	6837	5	41	19	19	91	15312	338	563	2	10	49	56	10	1028	16340
EXT 077	8050	6641	7	39	23	17	74	14851	309	478	2	9	53	51	9	911	15762
EXT 050	8156	6683	7	27	21	20	75	14989	274	459	0	10	53	58	15	869	15858
EXT 220	10070	8558	26	36	21	19	94	18824	362	585	1	2	152	115	14	1231	20055
EXT 181	9981	8487	27	30	19	16	84	18644	373	681	3	4	167	107	22	1357	20001

~14,106 + ~1066 = ~15,172

~16,500 in Ng S B, et al. *Nature* 461, 272-276 (2009).

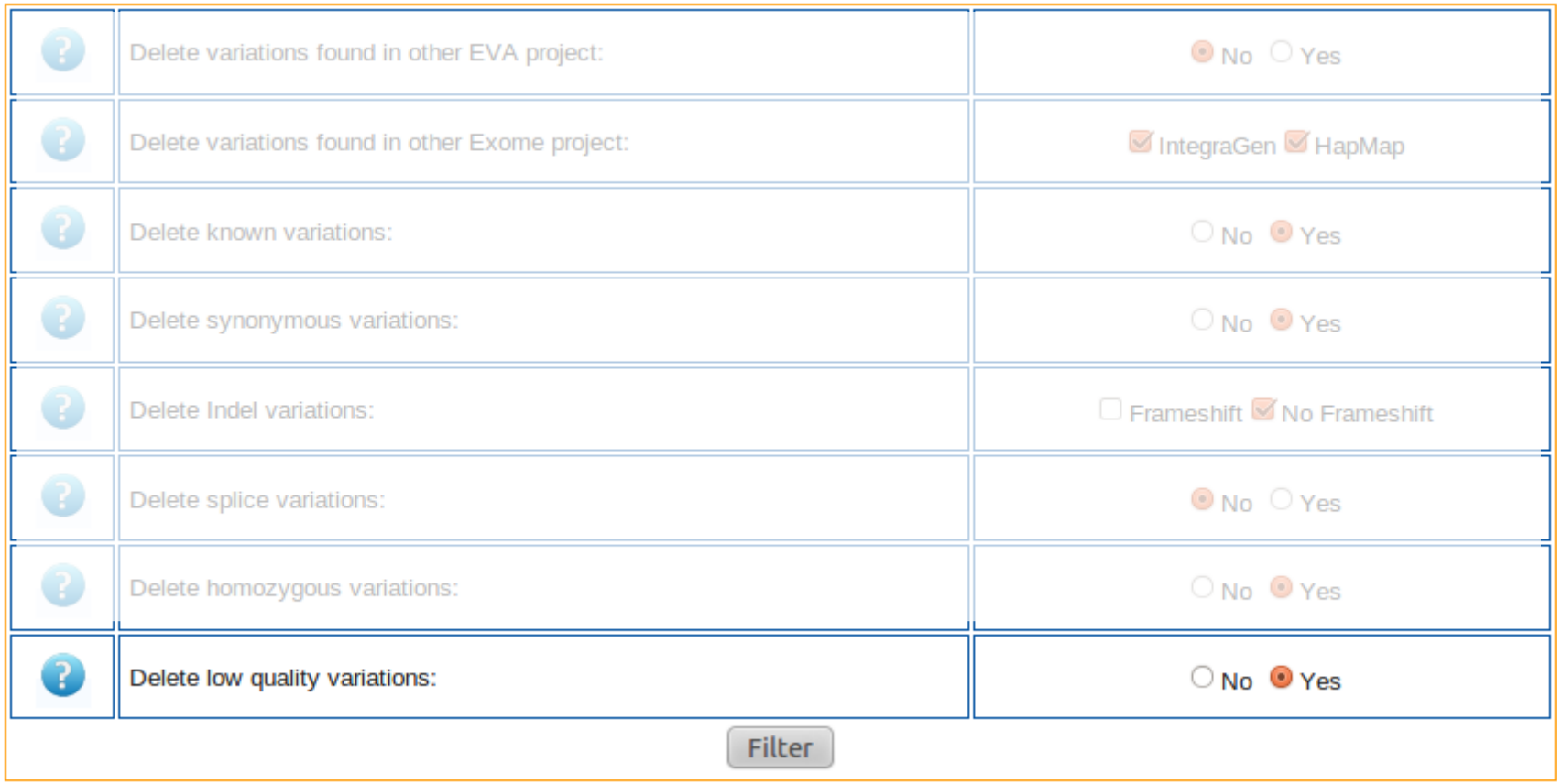
Recurrence Strategy - 2nd step: apply filters

Filters can be combined at will by clinicians to address different kinds of questions.
The combination is transformed into a SQL query and sent to the ExomeDB database.



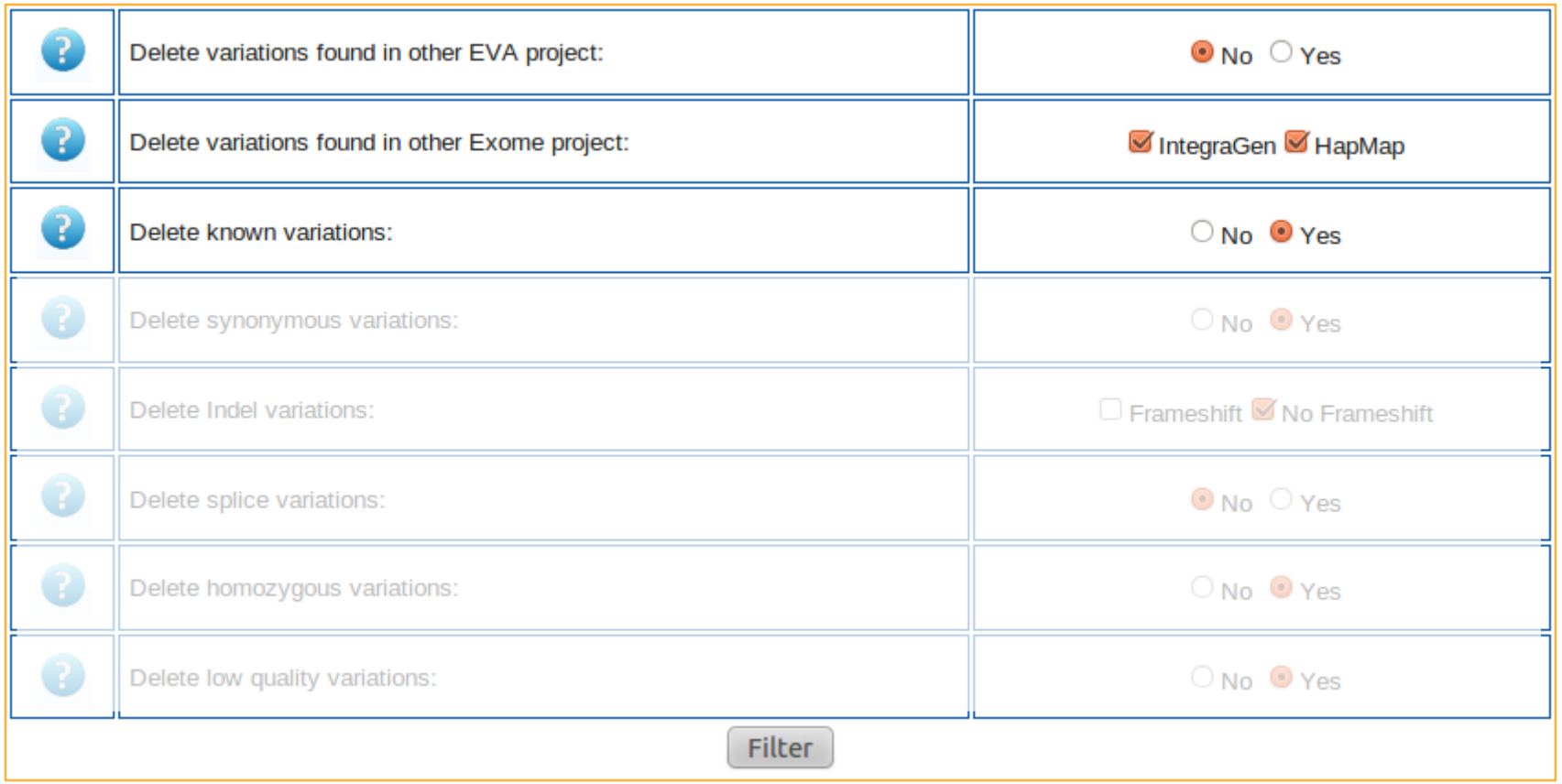
Filters
Recurrence Strategy - 2nd step: apply filters

Filters can be combined at will by clinicians to address different kinds of questions.
The combination is transformed into a SQL query and sent to the ExomeDB database.



Filters
Recurrence Strategy - 2nd step: apply filters

Filters can be combined at will by clinicians to address different kinds of questions.
The combination is transformed into a SQL query and sent to the ExomeDB database.



Filters
Recurrence Strategy - 2nd step: display filtered variations

[Variations overview]

Individual	Known variations								Unknown variations							
	Exon						Intron	Sub-total	Exon						Intron	Total
	Single variation				Indel		Splice		Single variation				Indel		Splice	
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2		Synonym	Missense	Stop loss	Nonsense	Fs	N Fs	+/- 2	
ALZ 049	0	0	0	0	0	0	0	**0**	0	286	0	8	25	0	3	322
ALZ 426	0	0	0	0	0	0	0	**0**	0	250	0	7	16	0	2	275
ROU 632	0	0	0	0	0	0	0	**0**	0	258	1	16	18	0	9	302
EXT 049	0	0	0	0	0	0	0	**0**	0	344	0	9	25	0	5	383
EXT 055	0	0	0	0	0	0	0	**0**	0	360	0	17	14	0	4	395
ALZ 062	0	0	0	0	0	0	0	**0**	0	328	1	9	20	0	2	360
ROU 816	0	0	0	0	0	0	0	**0**	0	336	0	9	22	0	5	372
ALZ 198	0	0	0	0	0	0	0	**0**	0	288	0	7	17	0	6	318
ALZ 056	0	0	0	0	0	0	0	**0**	0	237	1	10	3	0	4	255
EXT 094	0	0	0	0	0	0	0	**0**	0	266	0	5	4	0	2	277
EXT 077	0	0	0	0	0	0	0	**0**	0	208	1	6	8	0	4	227
EXT 050	0	0	0	0	0	0	0	**0**	0	210	0	10	9	0	5	234
EXT 220	0	0	0	0	0	0	0	**0**	0	394	0	1	31	0	7	433
EXT 181	0	0	0	0	0	0	0	**0**	0	435	1	0	29	0	12	477

Delete variations found in other EVA project:	○ No ○ Yes
Delete variations found in other Exome project:	☑ IntegraGen ☑ HapMap
Delete known variations:	○ No ◉ Yes
Delete synonymous variations:	○ No ◉ Yes
Delete Indel variations:	☐ Frameshift ☑ No Frameshift
Delete splice variations:	◉ No ○ Yes
Delete homozygous variations:	○ No ◉ Yes
Delete low quality variations:	○ No ◉ Yes

Filter

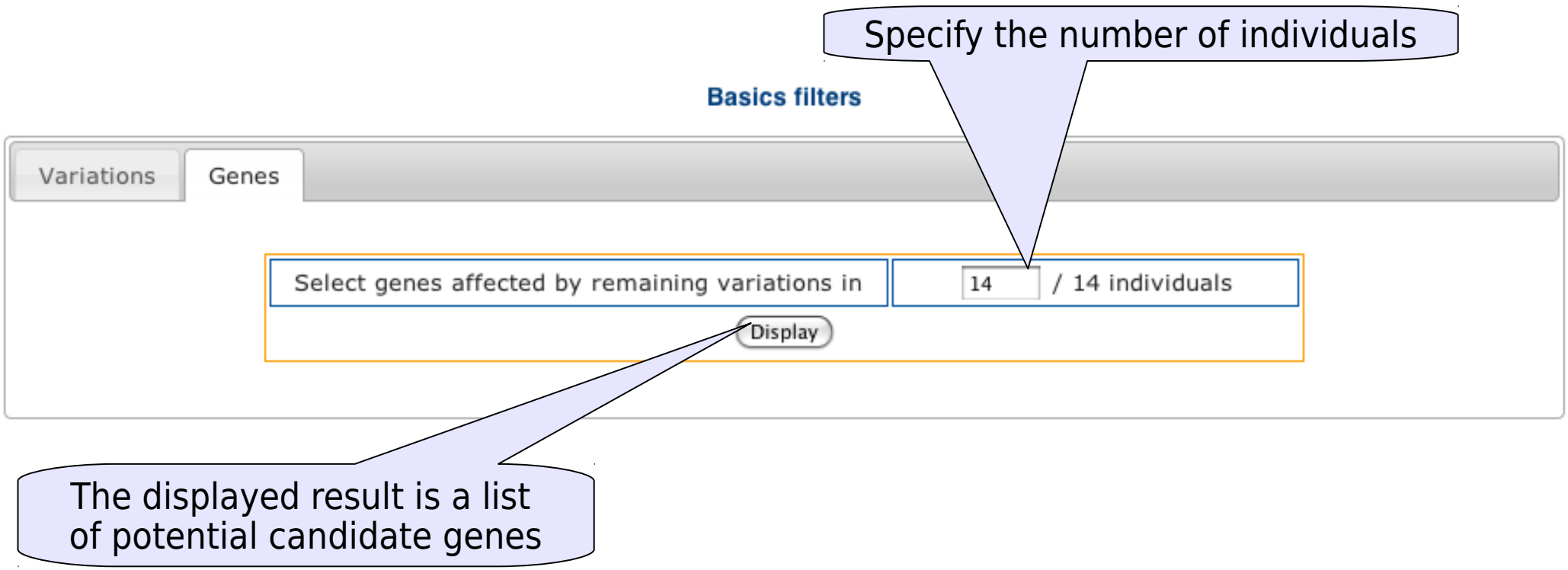
Only remain unknown, non synonym and high quality variations.

The number of variations per individual decreases from ~15,172 to ~330

Filters

Recurrence Strategy - 3rd step: retrieve corresponding gene & control the recurrence stringency

Search for the most affected gene in the specified number of individuals

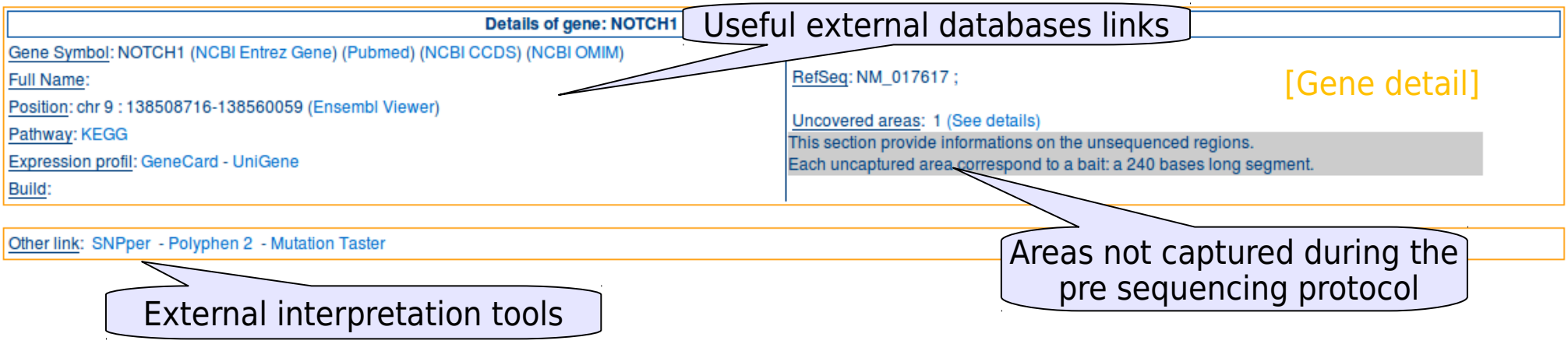


Results of recurrence strategy

The number of candidate genes drasticaly decreases with the number of individuals

Number of individuals	Number of genes with remaining variations
14 / 14	0
13 / 14	0
12 / 14	0
11 / 14	0
10 / 14	0
9 / 14	0
8 / 14	1
7 / 14	3
6 / 14	3
5 / 14	7
4 / 14	31
3 / 14	112
2 / 14	542
1 / 14	2730

Filters
Results - gene details, variations overview, variations list, variation details.



Filters
Results - gene details, variations overview, variations list, variation details.

Details of gene: NOTCH1

Useful external databases links

Gene Symbol: NOTCH1 (NCBI Entrez Gene) (Pubmed) (NCBI CCDS) (NCBI OMIM)
Full Name:
Position: chr 9 : 138508716-138560059 (Ensembl Viewer)
Pathway: KEGG
Expression profil: GeneCard - UniGene
Build:

RefSeq: NM_017617 ;

[Gene detail]

Uncovered areas: 1 (See details)
This section provide informations on the unsequenced regions.
Each uncaptured area correspond to a bait: a 240 bases long segment.

Areas not captured during the pre sequencing protocol

Other link: SNPper - Polyphen 2 - Mutation Taster

External interpretation tools

Affected individuals and detected variations

Numeric data

Variations overview for the gene

23 Known variations & 13 Unknown variations:

[Variations overview]

Individual	Known variations									Unknown variations								
	Exon						Intron	UTR		Exon						Intron	UTR	
	Single variation				Indel					Single variation				Indel				
	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs		5'	3'	Synonym	Missense	Stop loss	Nonsense	Fs	N Fs		5'	3'
AL 002	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
ALZ 6 010	2	0	0	0	0	0	4	0	0	0	0	0	0	0	0	1	0	0
RO 2 001	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
EXT 9 002	3	1	0	0	0	0	4	0	0	0	0	0	0	0	0	1	0	0
EXT 5 002	4	0	0	0	0	0	5	0	0	0	1	0	0	0	0	0	0	0
ALZ 2 001	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
RO 6 001	5	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
ALZ 8 008	3	0	0	0	0	0	8	0	0	0	1	0	0	0	0	1	0	0
ALZ 6 001	5	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
EXT 4 001	5	0	0	0	0	0	8	0	0	0	0	0	0	0	0	2	0	0
E 077	3	0	0	0	0	0	6	0	0	0	1	0	0	0	0	0	0	0
EX 001	1	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0
EXT 0 001	2	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0
E 181	4	0	0	0	0	0	9	0	0	0	1	0	0	0	0	0	0	0
TOTAL Project	8	2	0	0	0	0	13	0	0	1	7	0	0	0	0	5	0	0

Filters

Results - gene details, variations overview, variations list, variation details.

[Variations list]

Variation	Genomic Position	cDNA & aa Position	Functional class	Category	Codon.w -> Codon.m	aa.w -> aa.m	Other exomes frequencies het | hom	Gene symbol	Individuals	Genotype status
EVA.9.NOTCH1.E.15937	9:138531600	c.1501 atg.1500 p.500	Exon	Synonym	CTG -> CTA	L -> L	0 | 0	NOTCH1	A 049 002	het1
EVA.9.NOTCH1.E.51336	9:138520111	c.4059 atg.4058 p.1353	Exon	Missense	GGC -> GAC	G -> D	0 | 0	NOTCH1	EX 077 001	het1.D
EVA.9.NOTCH1.E.58998	9:138524932	c.2735 atg.2734 p.912	Exon	Missense	CGG -> TGG	R -> W	0 | 0	NOTCH1	E 050 001	het2
EVA.9.NOTCH1.E.58999	9:138532918	c.1046 atg.1045 p.349	Exon	Missense	ACC -> CCC	T -> P	8 | 0	NOTCH1	E 050 001	het1.D
EVA.9.NOTCH1.E.77894	9:138532461	c.1205 atg.1204 p.402	Exon	Missense	TCG -> CCG	S -> P	0 | 0	NOTCH1	R 632 001	het1
rs11574895	9:138524975	c.2692 atg.2691 p.897	Exon	Synonym	GCC -> GCT	A -> A	0 | 0	NOTCH1	EX 055 002 RC 816 001 AL 056 001	het2 het1 het2.D
rs2229971	9:138527753	c.2266 atg.2265 p.755	Exon	Synonym	AAT -> AAC	N -> N	42 | 11	NOTCH1	AL 426 010 EX 055 002 RC 816 001 AL 198 008 AL 056 001 EX 094 001 EX 077 001	het1 hom hom het1 het2 het1 het1

Filters

Results - gene details, variations overview, variations list, variation details.

rs2229971

[Variation details]

dbSNP: yes Direct link	HapMap: 7	IntegraGen: het: 42 | hom: 11

Position	Gene
Chromosome 9	NOTCH1
genomic: 138527753 cDNA: 2266 ATG: 2265 Protein: 755	strand: - NM_017617 Exon 14

Variation		
reference	mutation	
T	C	Exon
AAT	AAC	Synonym
N	N	

Coverage & Quality	
A: 12	Total: 22
C: 0	
G: 10	Score: 44.10
T: 0	
Used: 22	

SNP variation annotations (position, gene, mutation) and quality information.

Filters
Results - gene details, variations overview, variations list, variation details.

rs2229971

[Variation details]

dbSNP: yes Direct link	HapMap: 7	IntegraGen: het: 42 | hom: 11

Position	Gene
Chromosome 9	NOTCH1
genomic: 138527753 cDNA: 2266 ATG: 2265 Protein: 755	strand: - NM_017617 Exon 14

Variation		
reference	mutation	
T	C	Exon
AAT	AAC	Synonym
N	N	

Coverage & Quality	
A: 12	Total: 22
C: 0	
G: 10	Score: 44.10
T: 0	
Used: 22	

SNP variation annotations (position, gene, mutation) and quality information.

Filters

Results - gene details, variations overview, variations list, variation details.

rs2229971

[Variation details]

dbSNP: yes Direct link	HapMap: 7	IntegraGen: het: 42 | hom: 11

Position	Gene
Chromosome 9	NOTCH1
genomic: 138527753 cDNA: 2266 ATG: 2265 Protein: 755	strand: - NM_017617 Exon 14

Variation		
reference	mutation	
T	C	Exon
AAT	AAC	Synonym
N	N	

Coverage & Quality	
A: 12	Total: 22
C: 0	
G: 10	Score: 44.10
T: 0	
Used: 22	

SNP variation annotations (position, gene, mutation) and quality information.

Filters
Results - gene details, variations overview, variations list, variation details.

rs2229971

[Variation details]

dbSNP: yes Direct link	HapMap: 7	IntegraGen: het: 42 | hom: 11

Position	Gene
Chromosome 9	NOTCH1
genomic: 138527753 cDNA: 2266 ATG: 2265 Protein: 755	strand: - NM_017617 Exon 14

Variation		
reference	mutation	
T	C	Exon
AAT	AAC	Synonym
N	N	

Coverage & Quality	
A: 12	Total: 22
C: 0	
G: 10	Score: 44.10
T: 0	
Used: 22	

SNP variation annotations (position, gene, mutation) and quality information.

Filters
Results - gene details, variations overview, variations list, variation details.

rs2229971

[Variation details]

dbSNP: yes Direct link	HapMap: 7	IntegraGen: het: 42 | hom: 11

Position	Gene
Chromosome 9	NOTCH1
genomic: 138527753 cDNA: 2266 ATG: 2265 Protein: 755	strand: - NM_017617 Exon 14

Variation		
reference	mutation	Exon
T	C	
AAT	AAC	Synonym
N	N	

Coverage & Quality	
A: 12	Total: 22
C: 0	
G: 10	Score: 44.10
T: 0	
Used: 22	

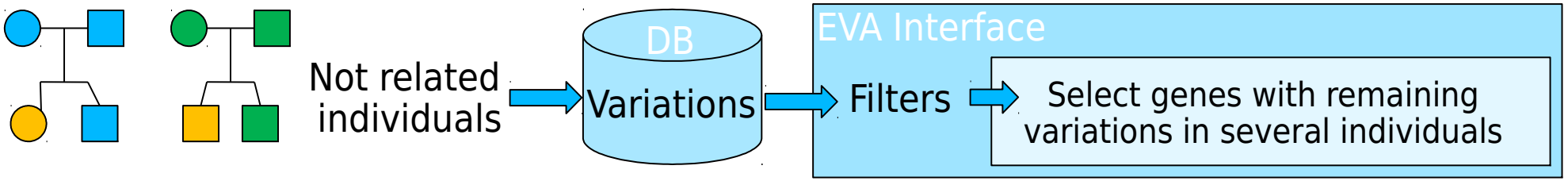
SNP variation annotations (position, gene, mutation) and quality information.

Other strategies: Familial and *de novo*

Filters: 3 ways of using them: recurrence familial and *de novo*

Recurrence strategy

Can be applied on dominant or recessive pathologies.

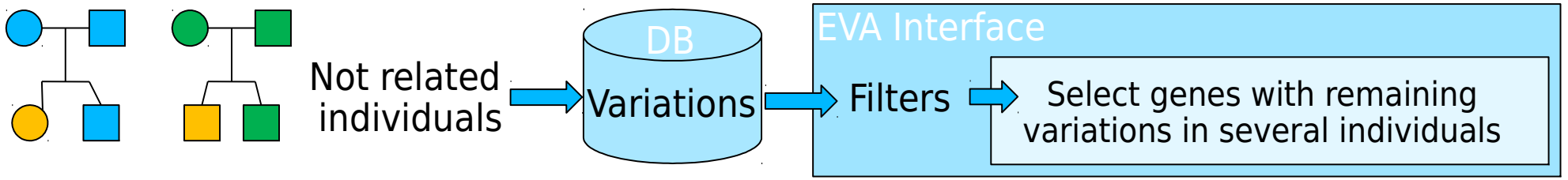


Other strategies: Familial and *de novo*

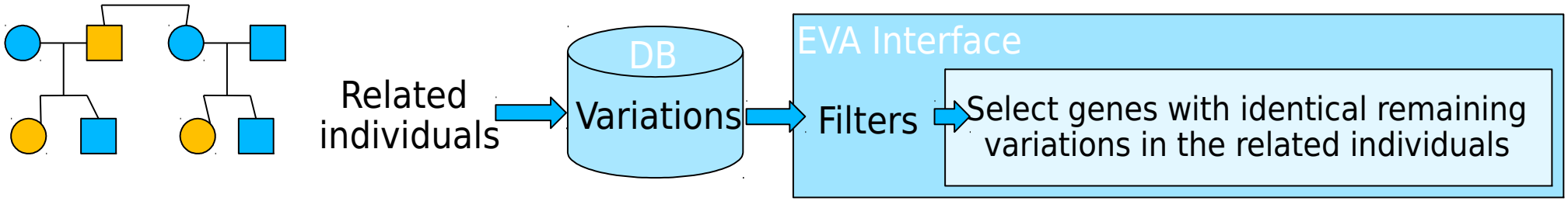
Filters: 3 ways of using them: recurrence familial and *de novo*

Recurrence strategy

Can be applied on dominant or recessive pathologies.



Familial strategy

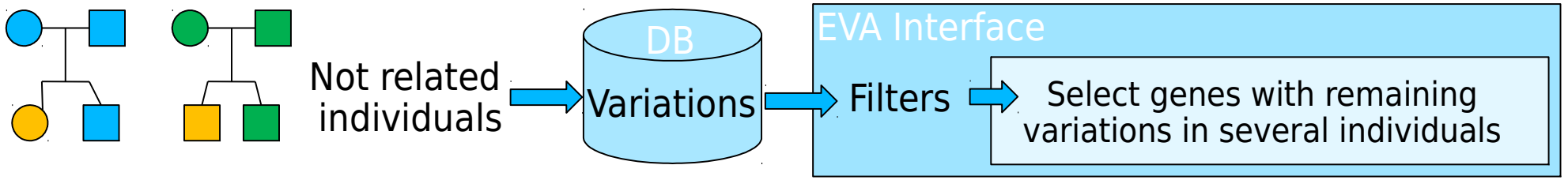


Other strategies: Familial and *de novo*

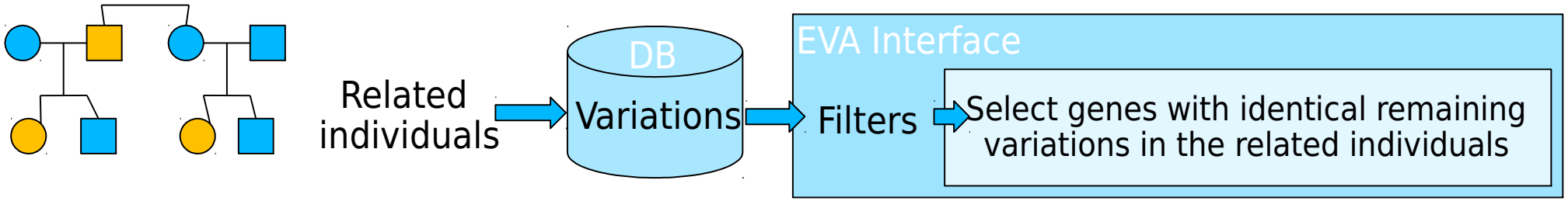
Filters: 3 ways of using them: recurrence familial and *de novo*

Recurrence strategy

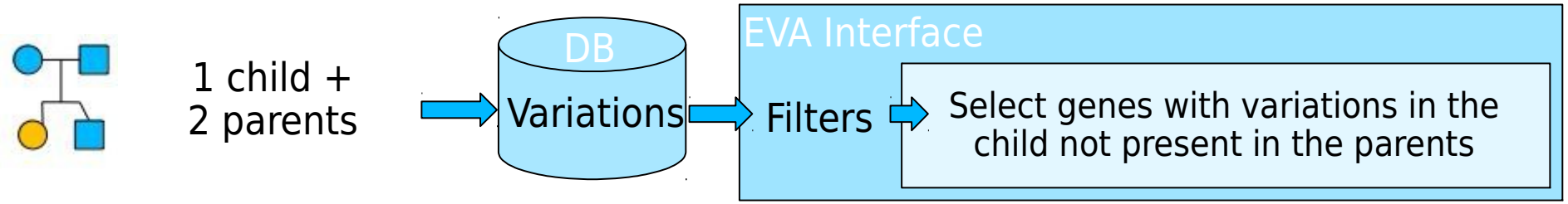
Can be applied on dominant or recessive pathologies.

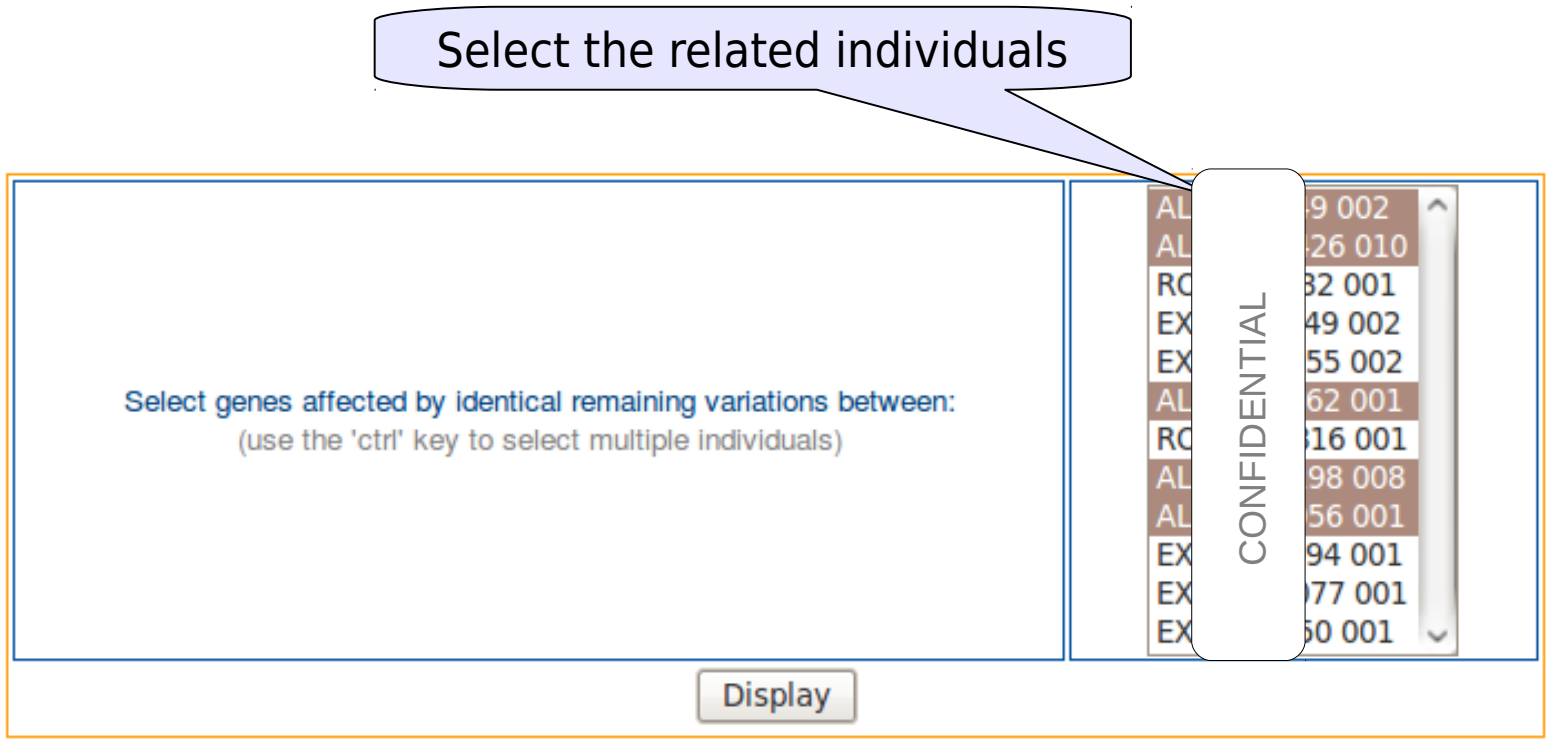


Familial strategy



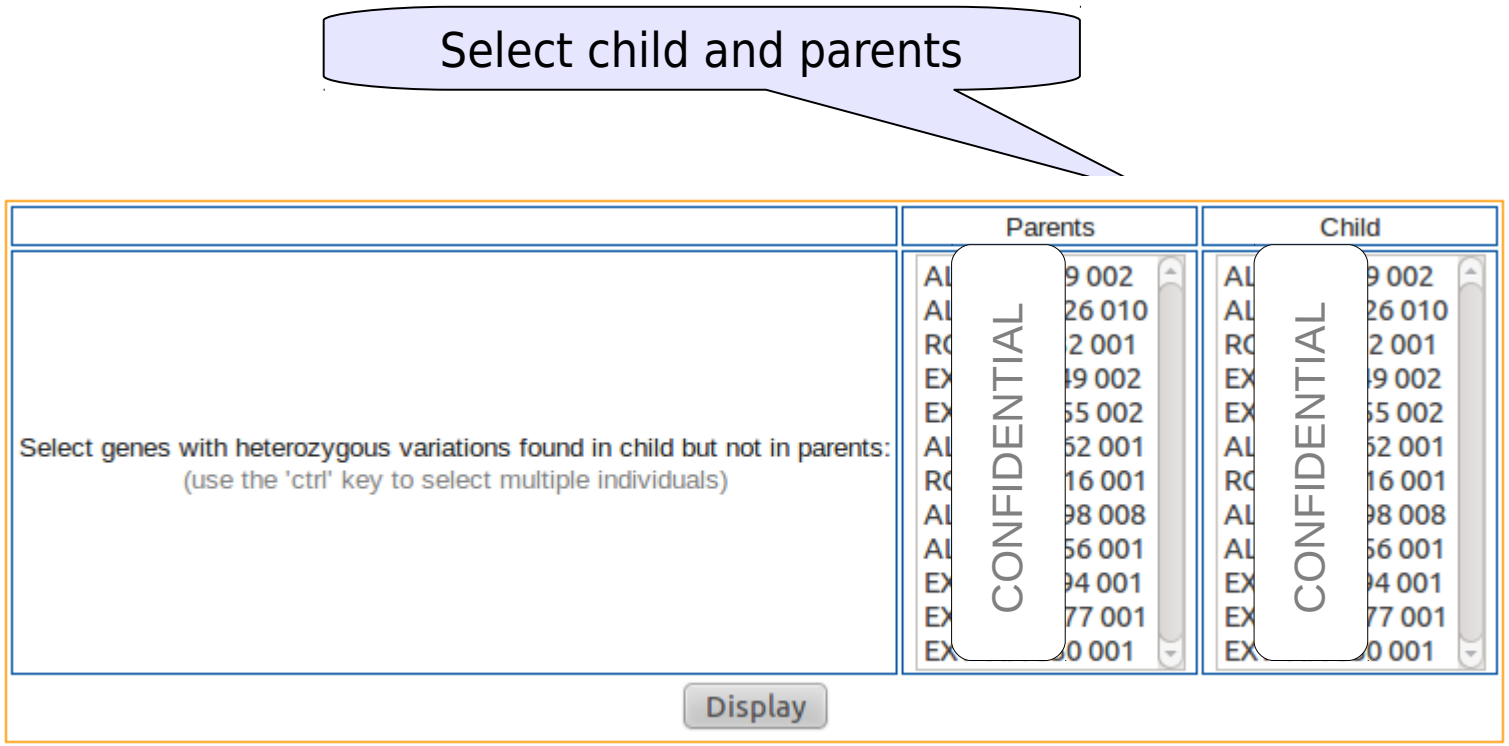
De novo strategy





Search for the gene affected by the most common variations
in the specified related individuals

Filters
De novo strategy



Search for the genes with remaining variations found in the child but not in the parents

Conclusion

EVA (Exome Variation Analyzer): simple, convivial and efficient tool.

- Database: ExomeDB
  store exome sequencing data

- Web interface:
  help clinicians in filtering and interpreting data

Results:

- Real decrease of candidate variations
- Case study (Alzheimer): 1 candidate gene revealed (publication submitted)

Encountered problems:

- Genes with frequent polymorphisms are not eliminated in Recurrence strategy
- Reference transcripts
- Variations found in other projects is necessary to drasticaly decrease the candidate list (1000 genomes, CompleteGenomics...)

Perspectives

Interface

Graphical representation

Tool

Integration module: VCF format compatibility

Statistical overview (project, individual, ...)

Availability

EVA is hosted on a dedicated server. The web address is public but an authentication must be given by an administrator in order to access biological data.

http://bioinfo.litislab.fr/EVA/

Thank you for your attention!