Development of an integrative approach for the derivation of signatures and translational analysis of cancer genomic data

***Georgia Kontogianni*[1,2], *Bianca Pasat*[2,3], *Eleftherios Pilalis*[2,3], *Olga Papadodima*[1], *Aristotelis Chatziioannou*[2,3]**

[1]*Metabolic Engineering and Bioinformatics Group, Institute of Chemical Biology, NHRF*
[2]*Center of Systems Biology, BRFAA*
[3]*e-NIOS Applications Private Company*

elixir GREECE

www.elixir-greece.org

This work:

We present a framework for streamlined analysis of NGS genomic data, integrating functional and pathway analyses, for the inference of gene signatures with diagnostic and/or prognostic value.

A system for the integration and processing of multi-layered data, implemented as a tool on the Galaxy platform.

A pilot analysis —including ten patients with cutaneous melanoma— resulting in a short list of candidate variants, mapped to genes with a probable pivotal role in the disease.

Our aim is to develop a tool for streamlined inference of gene signatures from NGS data allowing the accurate patient classification and clustering, towards a precision medicine approach.

What is melanoma?

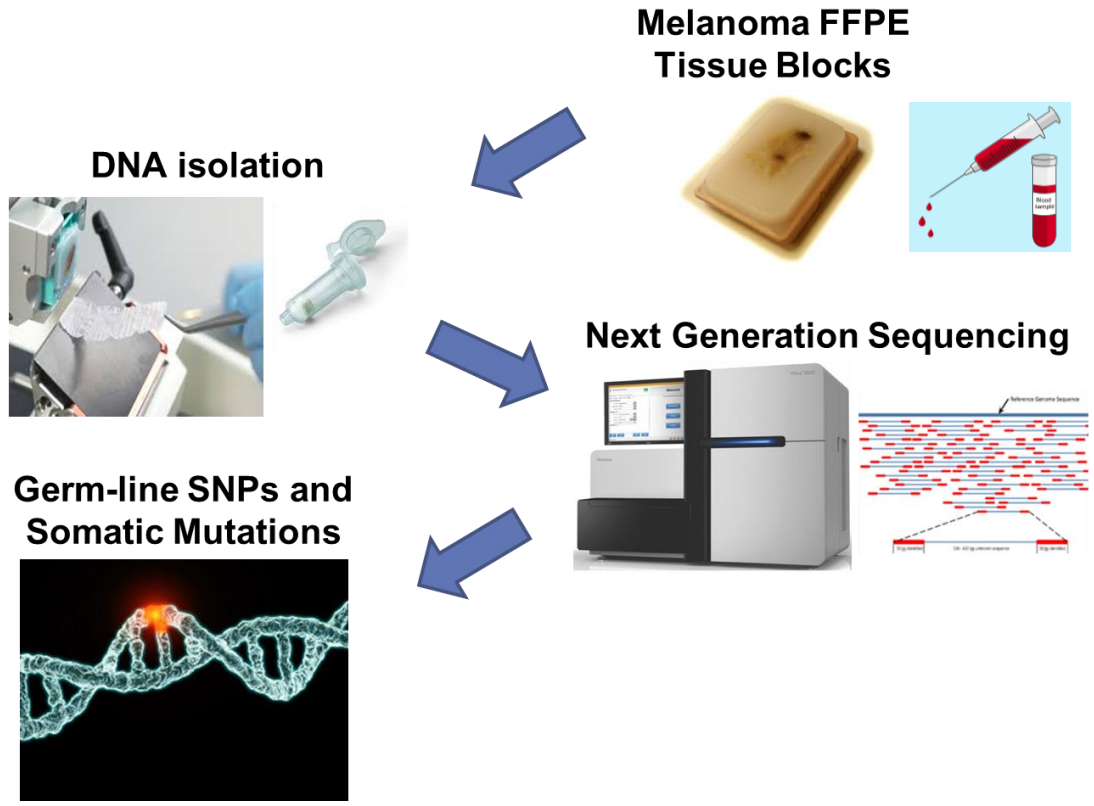
- Melanoma is a cancer that arises from epidermal melanocytes
- Accounts for less than 5% of skin cancer incidence, but it is responsible for the majority of skin cancer related deaths
- High mutation load
- UVR-characteristic signatures

The data:

Whole Exome Sequencing (WES)

10 patients with cutaneous melanoma

paired tumour tissue & blood



The workflow:

Raw sequencing data → **BWA** alignment → **Variant Calling** → **Biotranslator: Variant Analysis & Interpretation** Pathway analysis, signature inference

GATK
Variant discovery in high-throughput sequencing data

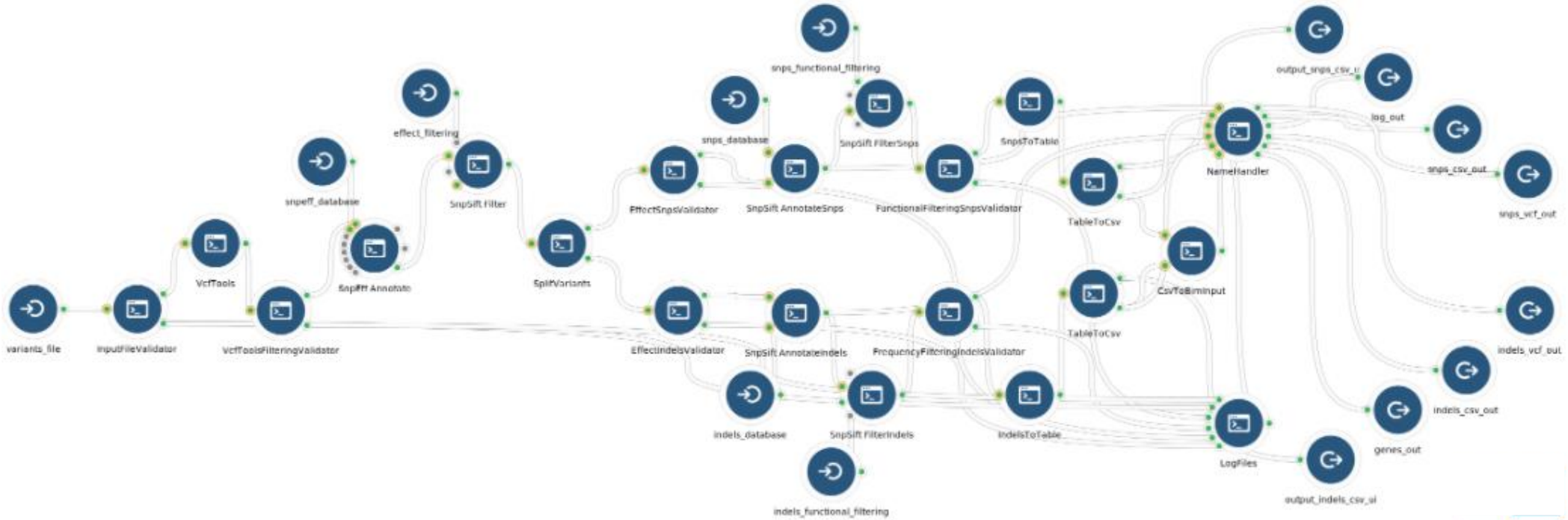
HaplotypeCaller
Germline variance calling

MuTect2
Somatic mutation calling

CNV workflow
Copy number variation calling

funcotator
Variance annotation

Biotranslator Variant Analysis:



tool for **annotating** and **filtering** variants from WES experiments
comprised of three main steps:

- pre-filtering (quality control metrics and thresholds),
- genomic annotation & filtering of genetic variants based on the effect prediction and
- custom filtering include functional and frequency filtering (available for SNPs).

Biotranslator Inputs & Outputs

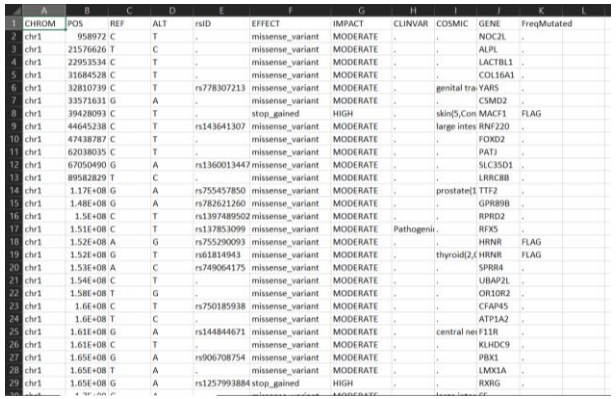
- **Variant Analysis**

Vcf file → **Variant annotation**

- **Interpretation**

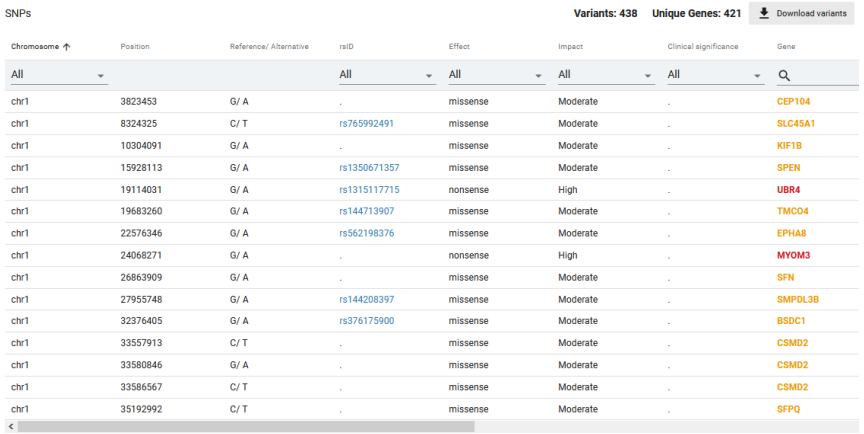
List of genes → **Functional analysis and prioritization**

Variant annotation



CHROM	POS	REF	ALT	rsID	EFFECT	IMPACT	CLINVAR	COSMIC	GENE	FreqMutated
chr1	958972	C	T	.	missense_variant	MODERATE	.	.	NOC2L	.
chr1	21576626	T	C	.	missense_variant	MODERATE	.	.	ALPL	.
chr1	22953534	C	T	.	missense_variant	MODERATE	.	.	LACTBL1	.
chr1	31684528	C	T	.	missense_variant	MODERATE	.	.	COL16A1	.
chr1	32810739	C	T	rs778307213	missense_variant	MODERATE	.	genital tra	YARS	.
chr1	33571631	G	A	.	missense_variant	MODERATE	.	.	C5MD2	.
chr1	39428093	C	T	.	stop_gained	HIGH	.	skin(5,Con	MACF1	FLAG
chr1	44645238	C	T	rs143641307	missense_variant	MODERATE	.	large intes	RNF220	.
chr1	47438787	C	T	.	missense_variant	MODERATE	.	.	FOXD2	.
chr1	62038035	C	T	.	missense_variant	MODERATE	.	.	PATJ	.
chr1	67050490	G	A	rs1360013447	missense_variant	MODERATE	.	.	SLC35D1	.
chr1	89582829	T	C	.	missense_variant	MODERATE	.	.	LRRC8B	.
chr1	1.17E+08	G	A	rs755457850	missense_variant	MODERATE	.	prostate(3	TTF2	.
chr1	1.48E+08	G	A	rs782621260	missense_variant	MODERATE	.	.	GPR89B	.
chr1	1.5E+08	C	T	rs1397489502	missense_variant	MODERATE	.	.	RPRD2	.
chr1	1.51E+08	C	T	rs137853099	missense_variant	MODERATE	Pathogenic	.	RFX5	.
chr1	1.52E+08	A	G	rs755290093	missense_variant	MODERATE	.	.	HRNR	FLAG
chr1	1.52E+08	G	T	rs61814943	missense_variant	MODERATE	.	thyroid(2,(	HRNR	FLAG
chr1	1.53E+08	A	C	rs749064175	missense_variant	MODERATE	.	.	SPRR4	.
chr1	1.54E+08	C	T	.	missense_variant	MODERATE	.	.	UBAP2L	.
chr1	1.58E+08	T	G	.	missense_variant	MODERATE	.	.	OR10R2	.
chr1	1.6E+08	C	T	rs750185938	missense_variant	MODERATE	.	.	CFAP45	.
chr1	1.6E+08	T	C	.	missense_variant	MODERATE	.	.	ATP1A2	.
chr1	1.61E+08	G	A	rs144844671	missense_variant	MODERATE	.	central ner	F11R	.
chr1	1.61E+08	C	T	.	missense_variant	MODERATE	.	.	KLHDC9	.
chr1	1.65E+08	G	A	rs906708754	missense_variant	MODERATE	.	.	PBX1	.
chr1	1.65E+08	T	A	.	missense_variant	MODERATE	.	.	LMX1A	.
chr1	1.65E+08	G	A	rs1257993884	stop_gained	HIGH	.	.	RXRG	.

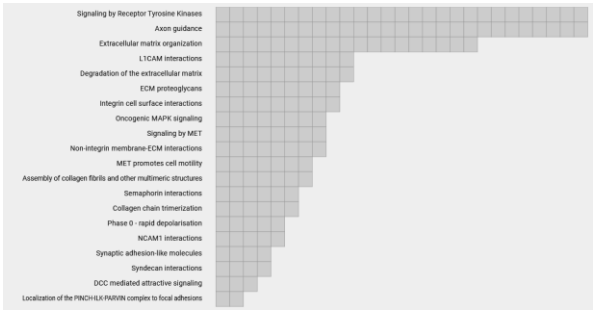
Variant filtering



SNPs — Variants: 438 — Unique Genes: 421 — Download variants

Chromosome	Position	Reference/ Alternative	rsID	Effect	Impact	Clinical significance	Gene
All			All	All	All	All	
chr1	3823453	G/ A	.	missense	Moderate	.	CEP104
chr1	8324325	C/ T	rs765992491	missense	Moderate	.	SLC45A1
chr1	10304091	G/ A	.	missense	Moderate	.	KIF1B
chr1	15928113	G/ A	rs1350671357	missense	Moderate	.	SPEN
chr1	19114031	G/ A	rs1315117715	nonsense	High	.	UBR4
chr1	19683260	G/ A	rs144713907	missense	Moderate	.	TMCO4
chr1	22576346	G/ A	rs562198376	missense	Moderate	.	EPHA8
chr1	24068271	G/ A	.	nonsense	High	.	MYOM3
chr1	26863909	G/ A	.	missense	Moderate	.	SFN
chr1	27955748	G/ A	rs144208397	missense	Moderate	.	SMPDL3B
chr1	32376405	G/ A	rs376175900	missense	Moderate	.	BSDC1
chr1	33557913	C/ T	.	missense	Moderate	.	CSMD2
chr1	33580846	G/ A	.	missense	Moderate	.	CSMD2
chr1	33586567	C/ T	.	missense	Moderate	.	CSMD2
chr1	35192992	C/ T	.	missense	Moderate	.	SFPQ

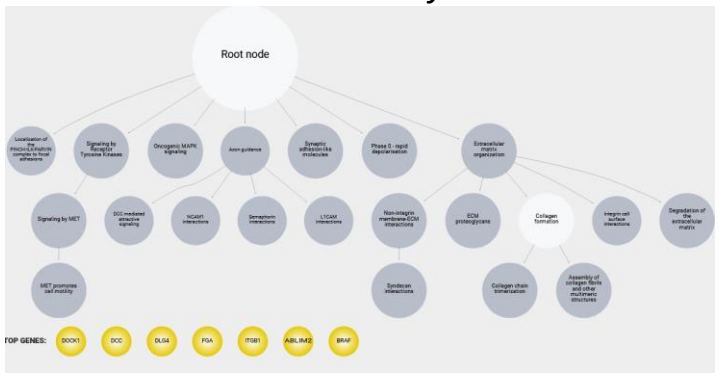
Pathway prioritization



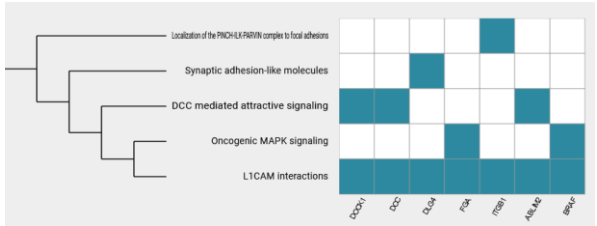
Gene/variant prioritization

Rank	Gene Symbol	Definition	Systemic P	Interactior	Drugs
1	NOTCH1	notch 1	8	7	0
2	ROCK1	Rho associated coiled-coil containing protein kinase 1	6	2	6
3	SFRP1	secreted frizzled related protein 1	5	0	0
4	SHOX2	short stature homeobox 2	5	0	0
5	WDR1	WD repeat domain 1	4	1	0
6	TGFBR2	transforming growth factor beta receptor 2	4	0	2
7	NLGN1	neuroligin 1	4	0	1
8	ACAN	aggrecan	4	0	1
9	VCAN	versican	4	1	1
10	GLI1	GLI family zinc finger 1	4	1	0
11	SLC1A3	solute carrier family 1 member 3	4	0	1
12	RAPGEF1	Rap guanine nucleotide exchange factor 1	3	0	0
13	AQP1	aquaporin 1 (Colton blood group)	3	0	1
14	CSPG4	chondroitin sulfate proteoglycan 4	3	0	0
15	KIF23	kinesin family member 23	3	2	0
16	LGR4	leucine rich repeat containing G protein-coupled receptor 4	3	8	0
17	MACF1	microtubule-actin crosslinking factor 1	3	0	0
18	KCNA1	potassium voltage-gated channel subfamily A member 1	3	0	8
19	ACVRL1	activin A receptor like type 1	3	0	2
20	FASN	fatty acid synthase	3	1	3
21	GOLGA5	golgin A5	2	1	0
22	SMYD3	SET and MYND domain containing 3	2	1	1

Functional analysis



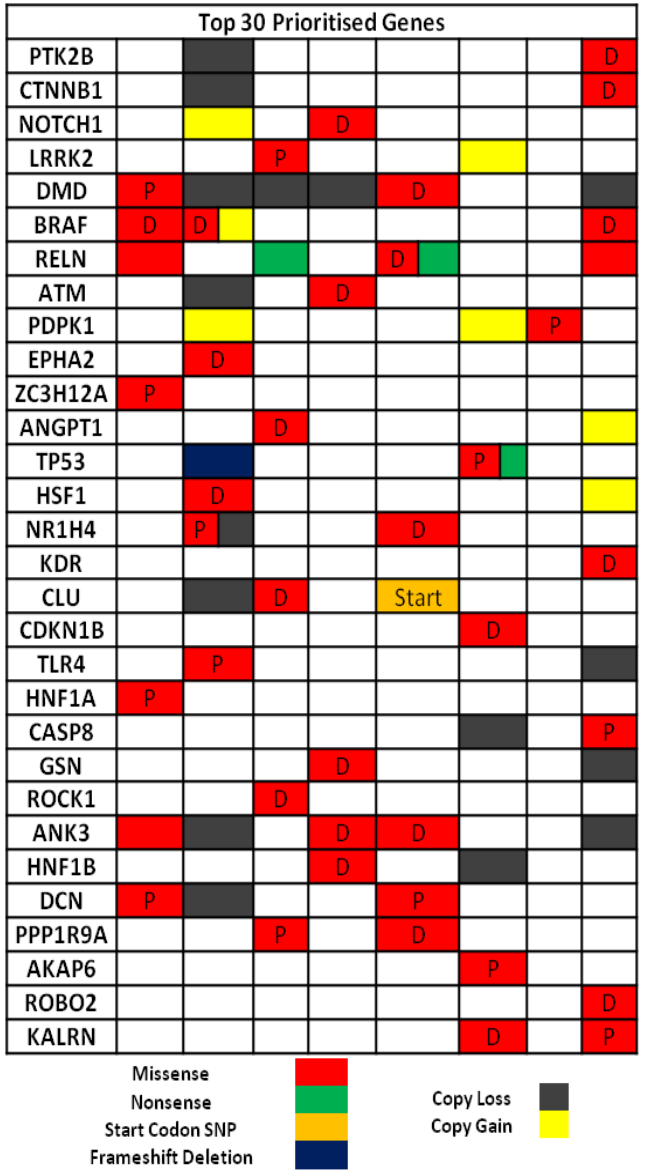
Gene signature



Ranking/ Signature inference in Melanoma

Biotranslator analysis results, based on their centrality (genes taking part in numerous distinct mechanisms are ranked higher), exploiting semantic information and network analysis

Top 30 prioritised genes, according to their network centrality, as described in Gene Ontology & Reactome vocabularies, and the mutations they carry in different patients



Discussion:

Analysis framework of NGS genomic data, integrating functional and pathway analyses

The tools will be interconnected using the galaxy platform, creation of workflow for reproducible analysis

Started adding the tools on Galaxy

Our analysis highlights a short list of candidate mutated genes with a probable pivotal role in melanoma that could be promising targets for future investigation

Acknowledgements

Georgia Kontogianni[1,2], Bianca Pasat[2,3], Eleftherios Pilalis[2,3], Olga Papadodima[1], Aristotelis Chatziioannou[2,3]

[1]*Metabolic Engineering and Bioinformatics Group, Institute of Chemical Biology, National Hellenic Research Foundation, Athens, Greece*

[2]*Center of Systems Biology, Biomedical Research Foundation of the Academy of Athens, Athens, Greece*

[3]*e-NIOS Applications Private Company, Kallithea, Greece*

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

We acknowledge support of this work by the project “TRANSITION: TRANSlating the complexIty of melanoma diagnosis into raTional therapeutic stratificatiON” (MIS 5031295) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

Co-financed by Greece and the European Union

We acknowledge support of this work by the project “ELIXIR-GR: Hellenic Research Infrastructure for the Management and Analysis of Data from the Biological Sciences” (MIS 5002780) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).