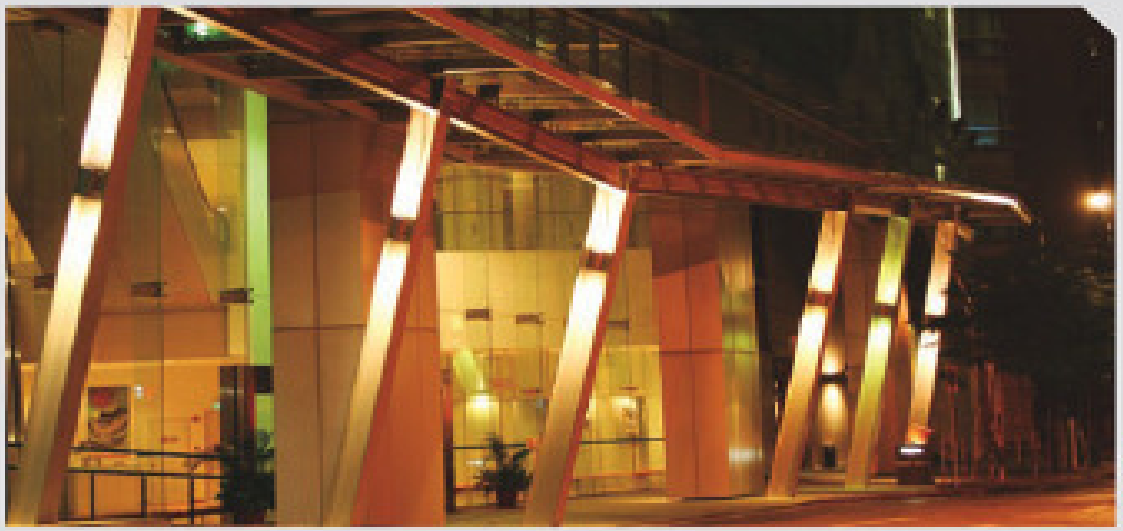
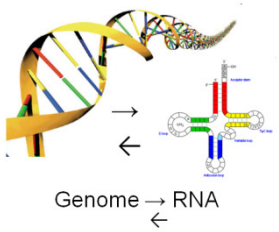


Gene Networks, tumor subtypes and patient prognogtic signatures associated with ovarian cancer mutations

Vladimir Kuznetsov

ECCB-2014



Division of Genome and Gene Expression Data Analysis

BII A*STAR

Integrative Genomics and Transcriptomics Analysis is bringing together many disciplines aiming understand and predict complex biological systems and eventually delivers solutions for Personalized Clinical Decision Support System and Clinical Data Management

- how to extract meaningful data from systems with ever incomplete and increasing complexity?
- how to analyse them in a way that allows new insights?
- how to generate data that is needed for clinical practice, but not yet available?
- how to find new empirical laws, or more fundamental theories, concerning how the complex systems work?
- how to personalize Big Data for a patient?

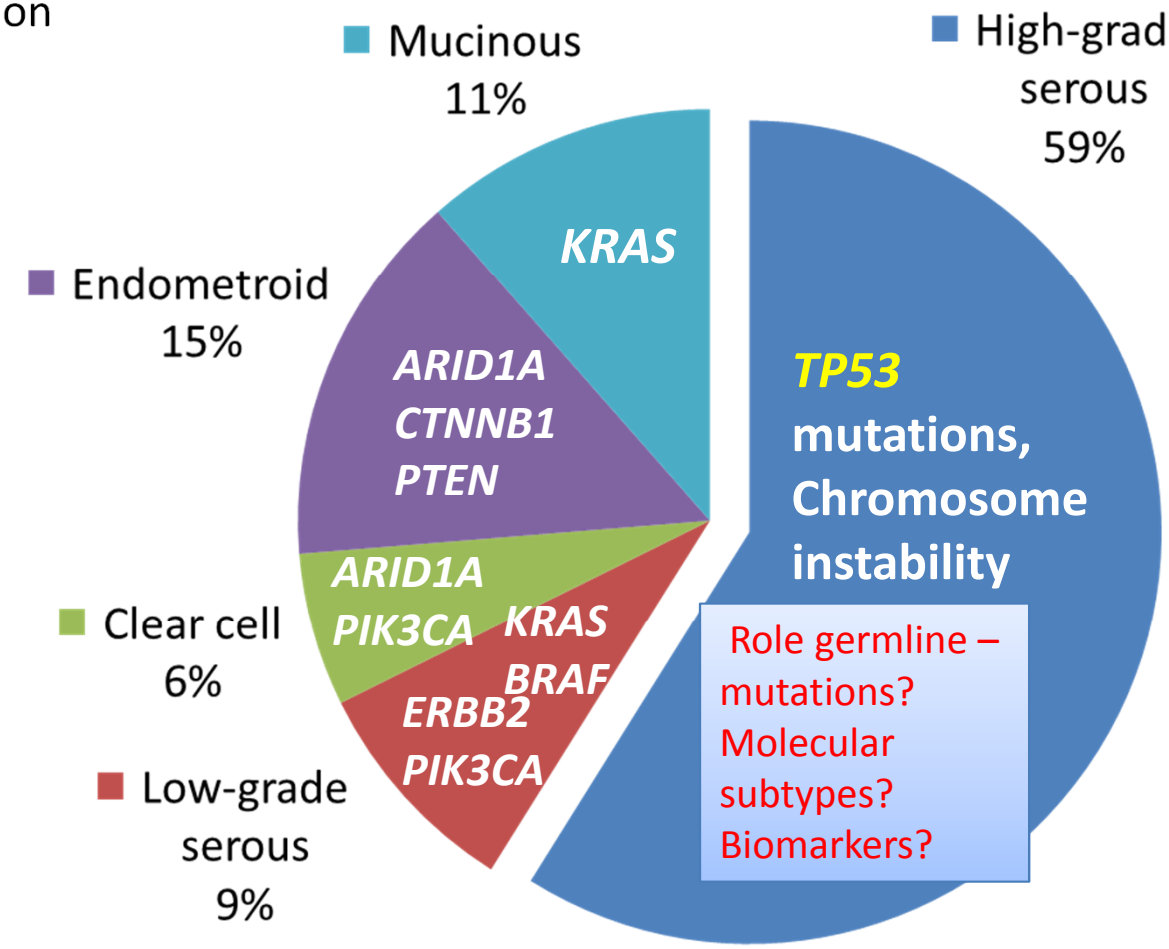
Cancers in the ovary: two or more types of the tumor genesis. Are there molecularly HG-SOC subtypes?

Type 1

Develop from cortical inclusion **cysts**

Type 2

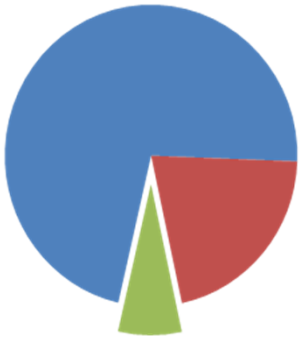
Originate in the **fallopian tubes**; migrate into the ovaries



Cancers in the ovary: the tumor origins

The fallopian tube is the main origin of high-grade serous ovarian cancer (Drapkin, Kurman, Herzog*)**:**

- Fallopian tube epithelium (Type 2) - 72%
- Ovarian surface epithelium (Type 1) - 21%
- Extragenital - 7%



Extragenital ovarian tumors originate mainly from primary gastrointestinal and breast tumors (Li W. et al**)

- Colon and rectum - 49%
- Stomach - 48%
- Breast - 8.2%
- Biliary duct - 1.4%
- Liver - 0.7%

T.J. Herzog, MD and H.E. Dinkelspiel, MD*. Curr Oncol. 2013 June; 20(3): 148–151

**Li W1, Wang H, Wang J, L V F, Zhu X, Wang Z. BMC Cancer. 2012 Jul 3;12:278. doi: 10.1186/1471-2407-12-278.

What actions can we take to improve the outcome for women with ovarian cancer(s)?

Bioinformatics Institute
A*STAR

Goals

- Improve molecular classification of OCs
- Discover the next generation clinical biomarkers for
  (i) tumor's classification and personalized
  (ii)early detection,
  (iii) prognosis,
  (iv) treatment prediction

REPORT

Cell Cycle 13:14, 2262–2280; July 15, 2014; © 2014 Landes Bioscience

Identification of two poorly prognosed ovarian carcinoma subtypes associated with *CHEK2* germ-line mutation and non-*CHEK2* somatic mutation gene signatures

Ghim Siong Ow[1], Anna V Ivshina[1], Gloria Fuentes[1,2], and Vladimir A Kuznetsov[1,3,4,*]

Ow Ghim Siong, Ivshina AV, Fuentes G, Kuznetsov VA.(2014) Identification of two poorly prognosed ovarian carcinoma subtypes associated with CHEK2 germ-line mutation and non-CHEK2 somatic mutation gene signatures. *Cell Cycle* (2014) May 30;13(14). [Epub ahead of print], published in the journal 15.07.2014

Bioinformatics Institute
A*STAR

Objectives

High-grade serous ovarian cancer(HG-SOC), a major histologic type of epithelial ovarian cancer(EOC), is a poorly-characterized, heterogeneous and lethal disease where somatic mutations of TP53 are common and inherited loss-of-function mutations in BRCA1/2 predispose to cancer in 9.5-13% of EOC patients. However, the overall burden of disease due to either inherited or sporadic mutations is not known.

Aims

Identification and characterization poor-prognosed serous ovarian carcinoma(SOC) subtypes driving by germline and somatic mutations

Data source

- The Cancer Genome Atlas (TCGA)
  - mRNA expression by U133A microarray
  - mRNA expression by RNA-seq
  - DNA Copy number
  - Mutation by AbiSolid and Illumina sequencing
  - Clinical information

Mutation data source

Exome sequencing of
- **334 high-grade serous ovarian carcinoma (HG-SOC)** tumor samples
  - analyzed via Illumina or ABI SOLID sequencing platforms
    - at the human genome sequencing centers (HGSCs)

↓

TCGA data portal (24th November 2010):
Level 2 data from BCM: 105 patients
Level 2 data from BI: 171 patients
Level 2 data from WUSM: 88 patients

↓

21,978 mutation data reported across 10,489 RefSeq gene symbols

↓

Removed 4,339 data where mutation status is Unknown or None.

17,639 mutation data reported across **9,083** RefSeq gene symbols

Data includes:
Mutation status: Germline, somatic or loss-of-heterozygosity
Variant classification: Missense_Mutation, Silent, Nonsense_Mutation, In_Frame_Del, Nonstop_Mutation, Splice_Site, Frame_Shift_Del, Frame_Shift_Ins, Splice_Site_Indel, In_Frame_Ins, Translation_Start_Site, Splice_Site_SNP or RNA
Variant type: Deletion, Insertion, SNP, DNP

A*STAR BII

Mutation data source (statistics)

Variant classification	Count	%
Missense_Mutation	11530	65.367
Nonsense_Mutation	737	4.178
Nonstop_Mutation	14	0.079
Frame_Shift_Del	413	2.341
Frame_Shift_Ins	243	1.378
In_Frame_Del	132	0.748
In_Frame_Ins	41	0.232
RNA	296	1.678
Splice_Site	374	2.120
Splice_Site_Indel	20	0.113
Splice_Site_SNP	14	0.079
Translation_Start_Site	1	0.006
Silent	3824	21.679
Grand Total	17639	100

Mutation status	Count	%
Somatic	**16458**	**93.3**
Germline	**970**	**5.5**
LOH	**211**	**1.2**
Grand Total	17639	**100**
Variant type	**Count**	**%**
SNP	**16717**	**94.8**
DEL	567	3.2
DNP	45	0.3
INS	310	1.8
Grand Total	17639	100

BII
A*STAR

Data collecting table

gene-patient tumor sample mutation frequency matrix

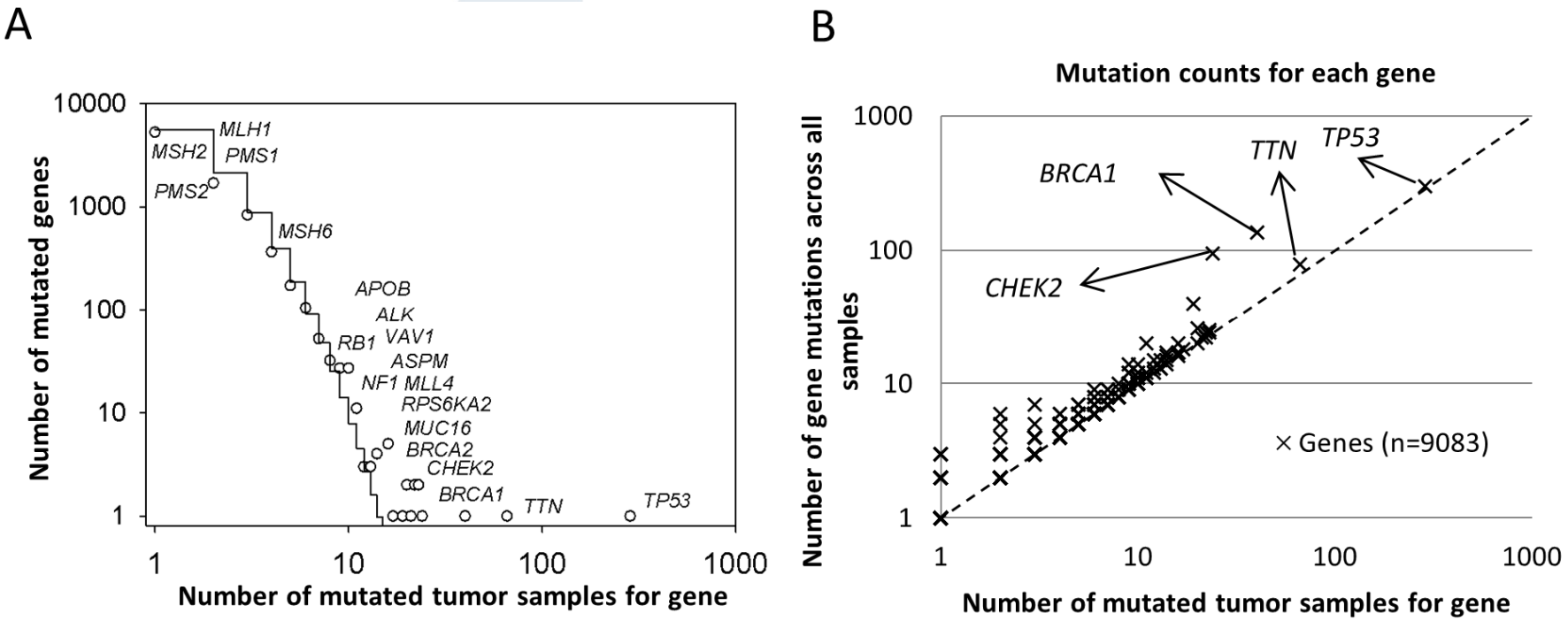
		Patients										sum
		p1	p2	p3	p4	p5	p6	p7	…	…	…	pM	
Genes	g1						x						1
	g2	x		x									2
	g3		x	x									2
	g4									x			1
	g5	x	x	x	x	x	x	x	x		x	x	10
	g6		x										1
	g7		x					x		x			3
	…												…
	…												…
	…												…
	gN			x	x					x			3

Two-dimensional gene-patient tumor sample mutation frequency matrix where the rows and columns correspond to 9083 unique gene symbols and 334 unique tumor sample IDs respectively.

Contingency table: # mutated genes for a given patient vs # mutated sample for a given gene

	#Mutated genes for a given patient	63	142	129	46	6	19	51	89	35
	#Mutated genome locations for a given patient across all genes	65	145	130	46	6	19	51	89	35
#Mutated genome locations for a given gene across all patients	Gene	TCGA-04-1337	TCGA-04-1338	TCGA-04-1347	TCGA-04-1348	TCGA-04-1351	TCGA-09-1666	TCGA-09-1669	TCGA-09-2044	TCGA-09-2045
1	A1CF	0	0	0	0	0	0	0	0	0
1	A2BP1	0	0	0	0	0	0	0	0	0
1	A2M	0	1	0	0	0	0	0	0	0
4	A2ML1	0	0	0	0	0	0	0	0	1
1	AAAS	0	0	0	0	0	0	0	0	0
2	AACS	0	0	0	0	0	0	0	0	0

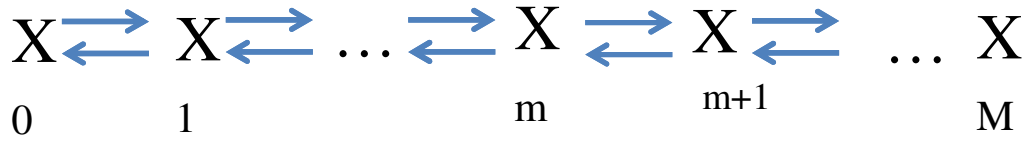
Number of distinct mutations vs number of mutated samples



- The frequency distribution of the number of mutated tumor samples(patients) is skewed and is fitted well by our Kolmogorov-Waring probabilistic function. This function predicted that only 31% of entire population cases were detected.
- HG-SOC is typically characterized by TP53 mutations but on average, only one TP53 mutation was observed per patient.
- On the other hand, CHEK2 and BRCA1 were observed to be mutated in less patients, but they can be mutated more frequently in the same genes of the given patient

Statistic modeling of mutation-repair process in cancer : Kolmogorov-Waring random process (Kuznetsov, 2003)

m: a random number of the patients for a given mutated gene follows to linear Markov birth-death process



Poisson + preferential occurrence

The probability of occurrence of m+1 samples(patients) for a given mutated gene

$$p_{m+1} = q\frac{(a+m)}{b+m+1}p_m$$

The probability of non-observed mutated genes (b>a>0)

$$p_0 = (1-\frac{a}{b})$$ (a: rate of mutation; b: rate of reparation)

By using curve fitting algorithm (Kuznetsov, 2003) applied to the observed frequency distribution data, the parameters were estimated

a=3.94; b=9.5; q=0.867

1. Known Mutations of SOC tumors and its Statistics

Published mutations associated with HG SOC:

- [1] MUC16, TP53, NF1, RB1, FAT3, CSMD3, GABRA6, CDK12, BRCA1,BRCA2;
- [2]+ SMARCB1, KRAS, NRAS, CREBBP, ERBB2
- [3] + BRIP, **CHEK2**, MRE11A, MSH6, NBN, PALB2, RAD50 and RAD51C

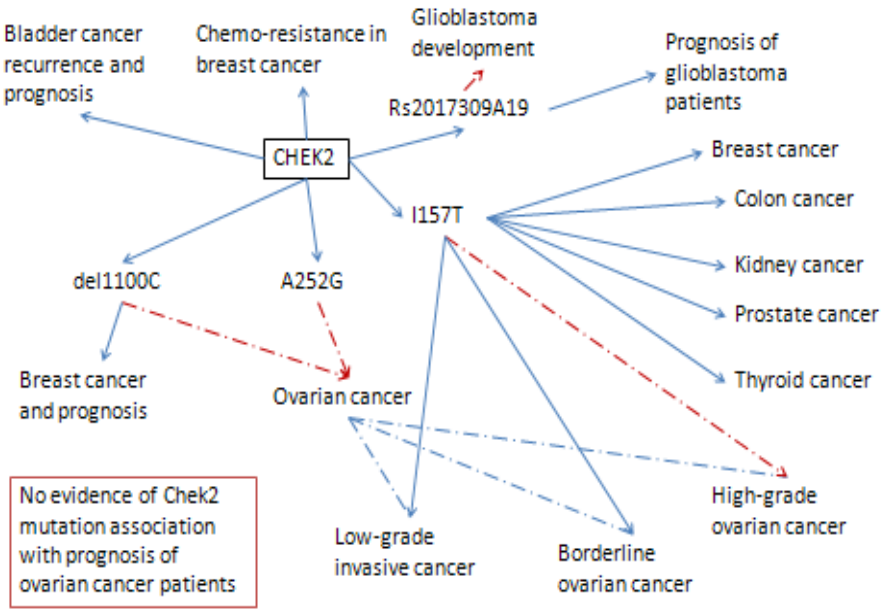
References:

[1] TCGA Research *Nature* 2011; 474:609-15.

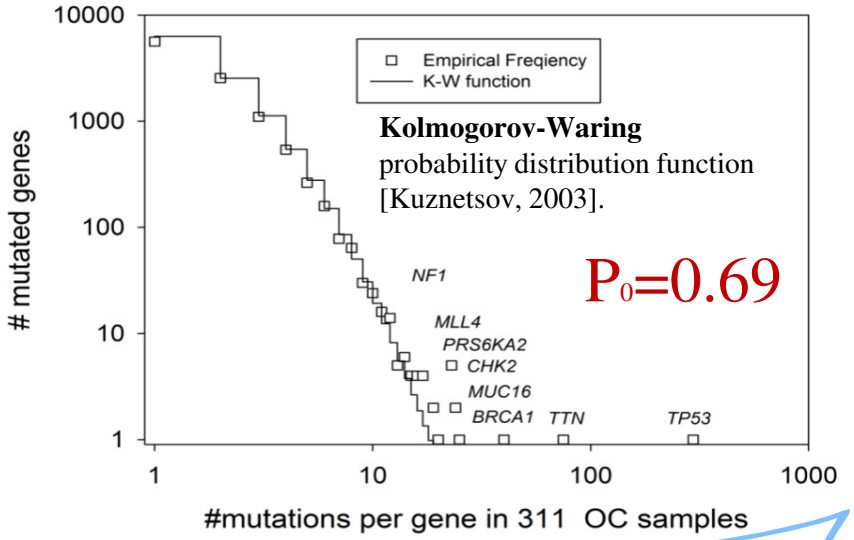
[2] Lawrence et al. *Nature* 2014; 505:495-501.

[3] Walsh et al. *PNAS USA* 2011; 108:18032-7.

CHEK2 and disease association



Frequency distribution of the mutations per gene in 311 TCGA HG-SOC samples. N=11,169



I. Is the list of **HG-SOC mutated genes** complete?

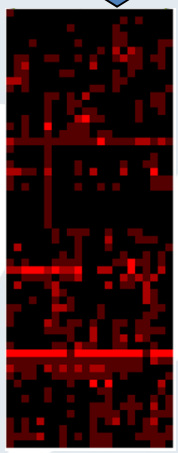
II. What was lost? (many rare mutations)

III. Could we used TCGA mutation data for prediction of novel classes of **HG-SOC?**

IV. Is there a benefit from mutation data for the patient prognosis and treatment prediction?

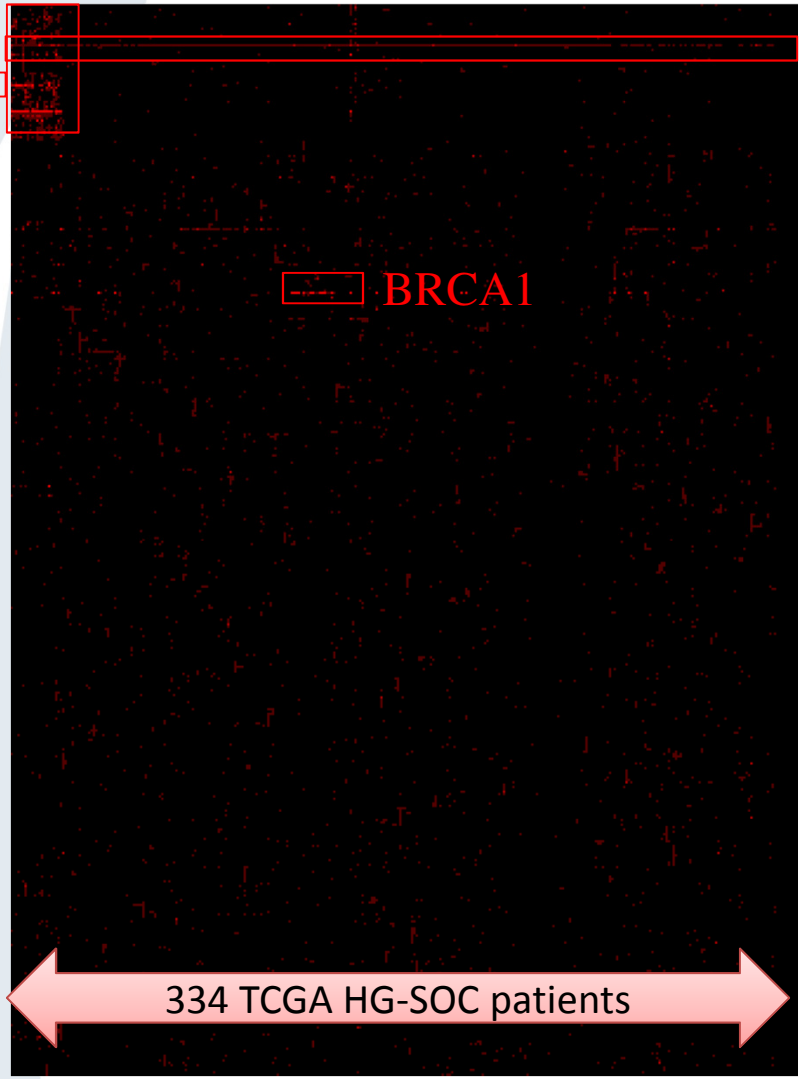
Hierarchical clustering of gene-patient mutation matrix revealed a subgroup of genes and patients with high mutation intensity

A high-intensity sub-cluster belonging to 22 patients (columns) and 58 genes (rows).



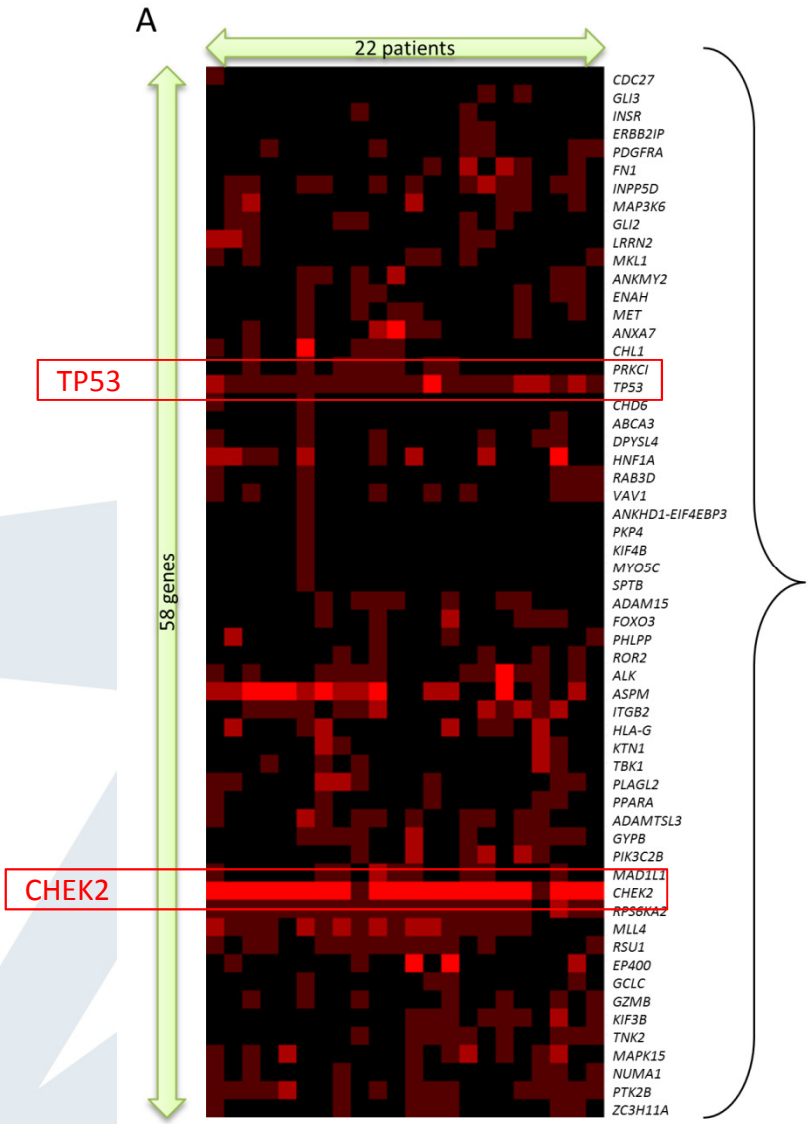
CHEK2

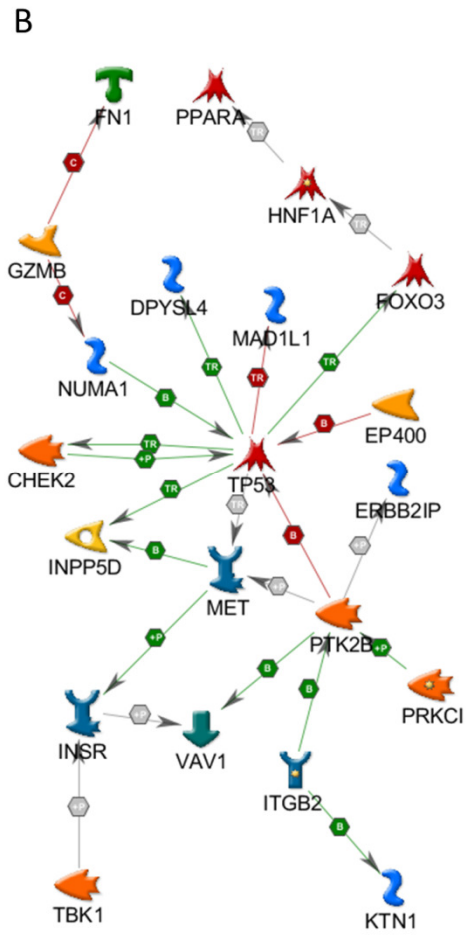
Next slide



BII
A*STAR

A subgroup of HG-SOC patients is characterized by high frequency of mutations in CHEK2





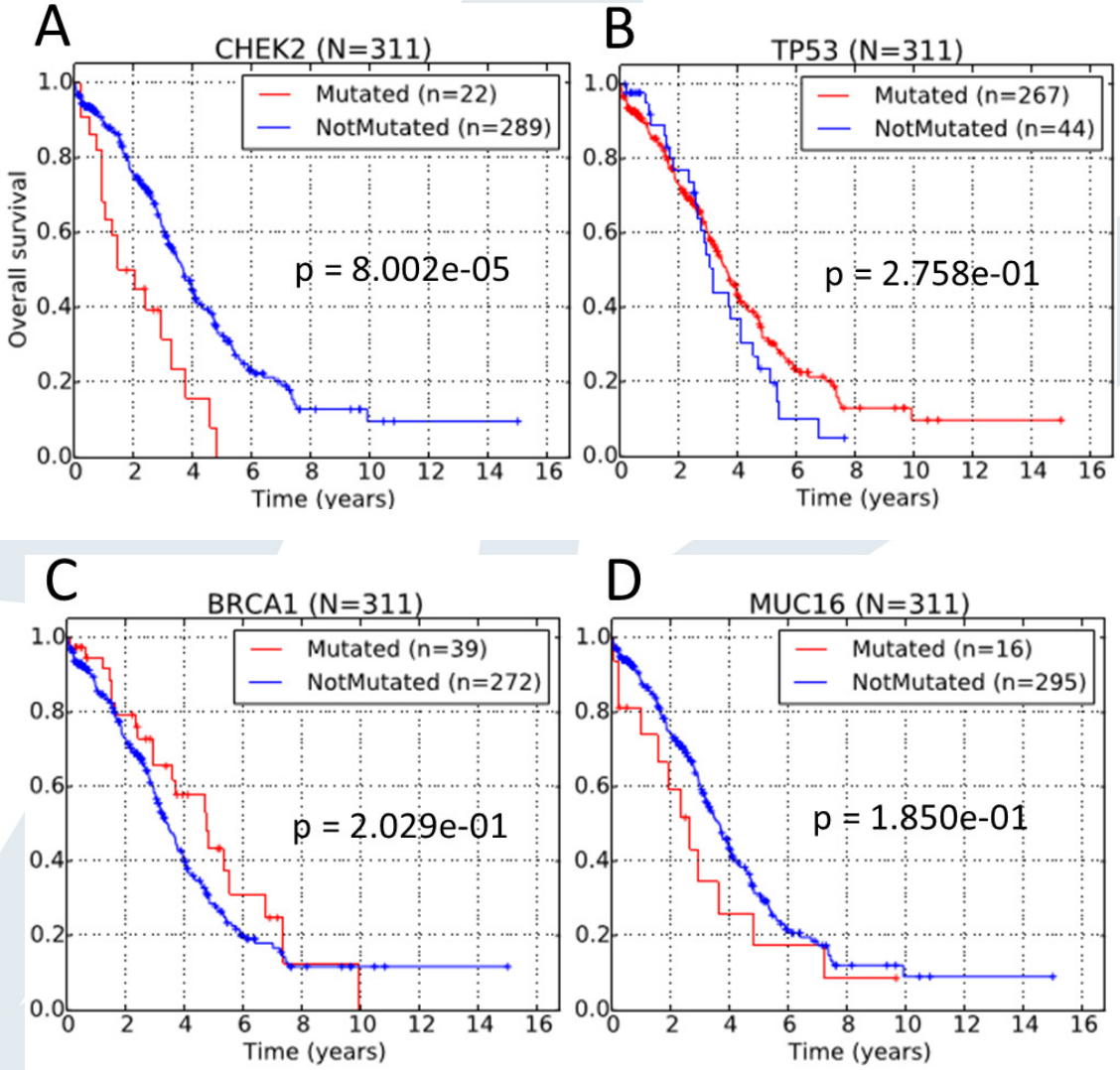
GO analysis of 58 genes revealed that these genes are significantly enriched in:

- Protein kinase activity
- Purine ribonucleotide binding
- Disease mutations

Direct interacting network of genes mostly involved in apoptosis, cell cycle control, DNA damage response and immune response.

BII
A*STAR

Patients with observed non-silent CHEK2 mutations exhibited significantly poorer overall survival rates



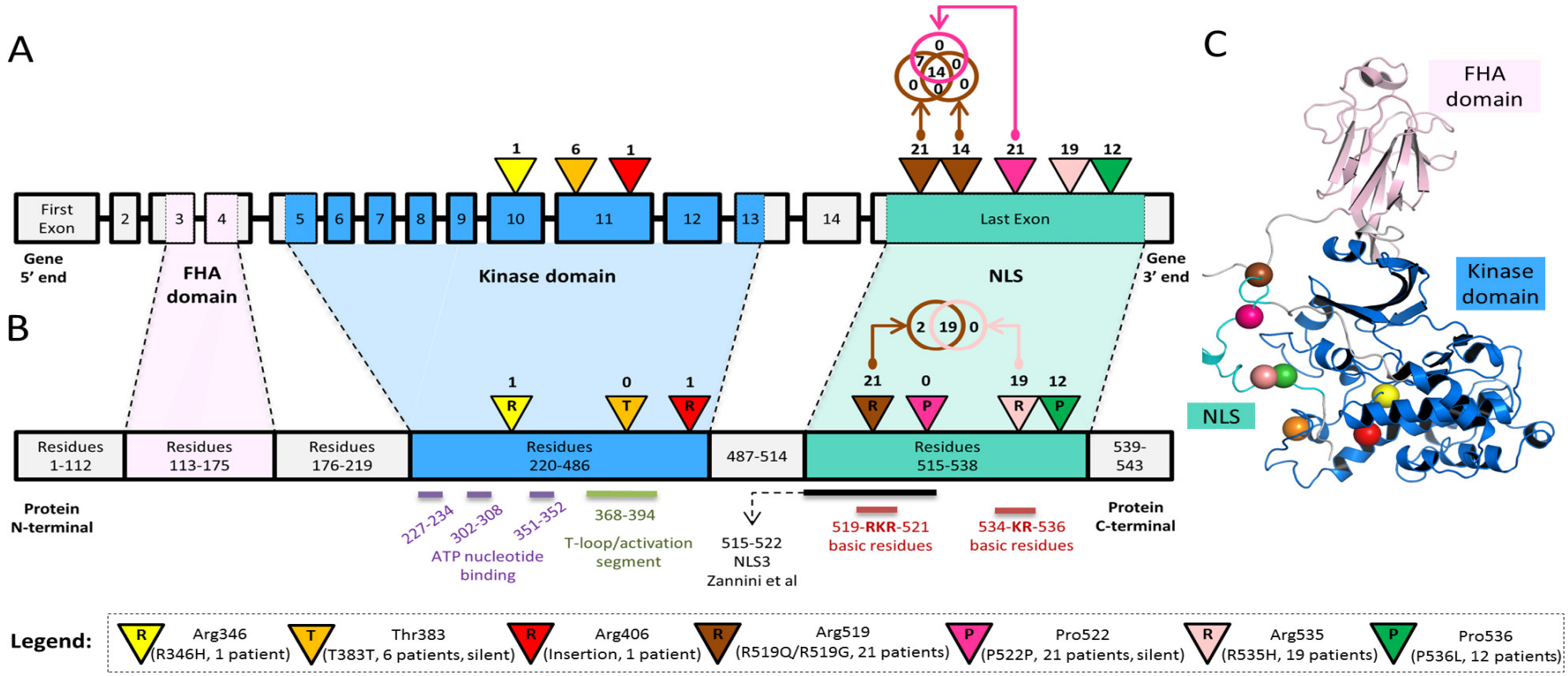
CHEK2 mutations are strongly correlated with poor patient overall survival.

TP53 and BRCA1 were mutated more frequently (in more patients), but they were not effective as prognostic biomarker.

MUC16 was also not very effective in patient prognosis.

A*STAR BII

Locations of mutation sites in CHEK2

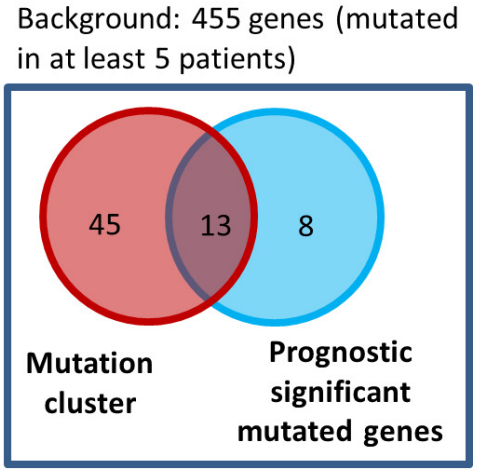


Mutations of the CHEK2 locus mostly occurred at the nuclear localization signal within the last exon, potentially implicating the nuclear import of the Chk2 protein and subsequent downstream biological processes. The coloured spheres indicate the locations of Chk2 mutations, which could be useful as a prognostic marker for HG-SOC. The forkhead-associated (FHA) domain, kinase domain and **nuclear localization signal (NLS)** are marked in pink, blue and cyan, respectively. Folded structures: Beta-strand (arrows); alpha helixes (helixes).

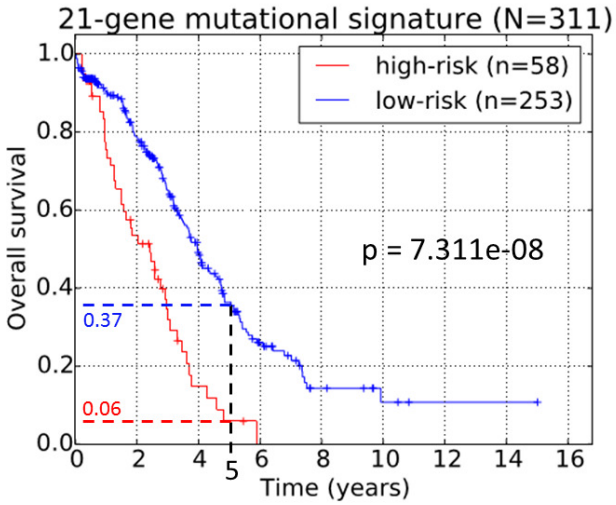
Identification of prognostic signature based on mutational status

ADAMTSL3
RPS6KA2
CHEK2
MLL4
RSU1
PTK2B
TNK2
GYPB
GLI2
???

A



B



A) There is considerable overlap between the genes from the mutation cluster and signature genes.

B) Mutational signature can stratify HG-SOC patients into high (5-year overall survival rate of 6%) or low-risk (5-y OS of 37%) subgroups.

BII
A*STAR

Mutational status of CHEK2 and the prognostic signature significantly correlates with poor response to therapy.

Prognostic marker or signature	Classification	Progressive disease	Complete or partial response or stable disease	Kappa (p-value)
(A) CHEK2	With mutation	5	15	0.1422 (0.03769)
	Without mutation	21	217	
(B) 21-gene signature	High-risk	8	42	0.08984 (0.06065)
	Low-risk	18	190	

Table: Kappa correlation of patients classified by the mutation status of (A) CHEK2 and (B) 21 gene mutational signature with therapy resistance.

(A)
25% of patients (5 of 20) with non-silent CHEK2 mutations exhibit progressive disease.
Only 8.8% of patients (21 of 237) without CHEK2 mutations exhibit progressive disease.

(B)
15% of patients (8 of 50) classified as high-risk exhibit progressive disease.
Only 8.7% of patients (18 of 208) classified as low-risk exhibit progressive disease.

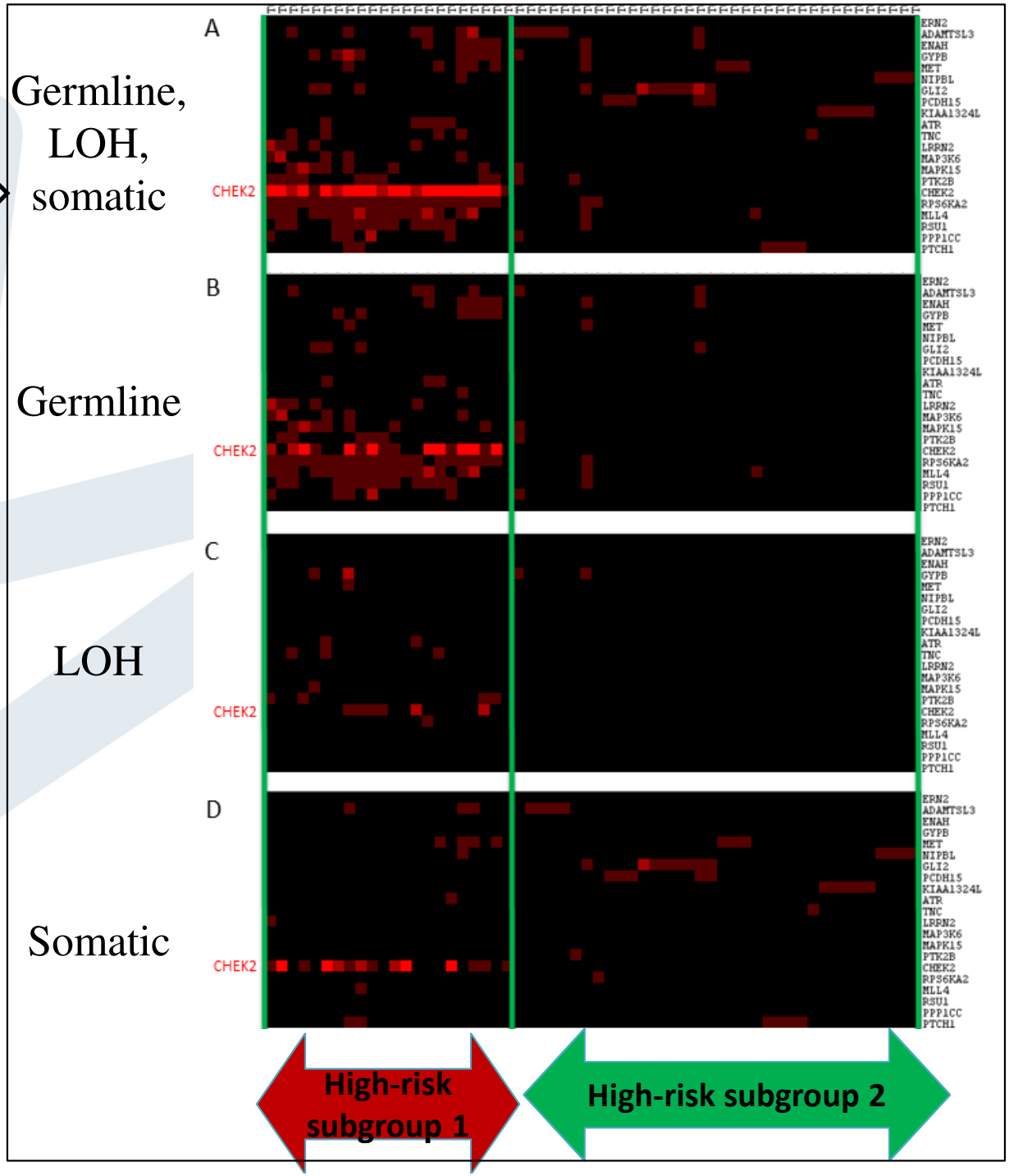
A*STAR BII

Heatmap clusters of non-silent mutations of the 21 prognostic genes and 58 patients in the poor prognosis subgroup:

We identified two tumor subclasses from the signature-defined high-risk subgroup:

High-risk subgroup 1:
- Characterized by **germline** mutations of CHEK2

High-risk subgroup 2:
- Characterized by **somatic** mutations of other prognostic genes



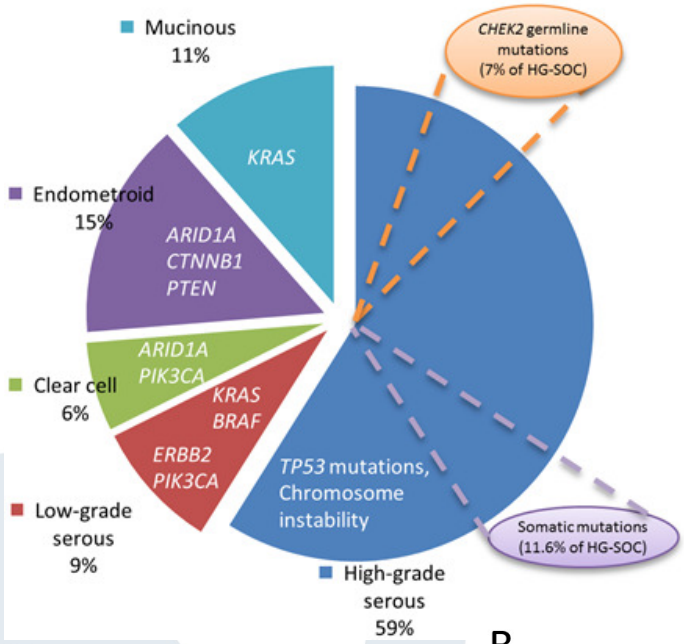
Co-occurrence of mutations was observed in CHEK2, RPS6KA2 and MLL4

Sample ID (n=22)	CHEK2	RPS6KA2	MLL4	CHEK2 and/or RPS6KA1 and/or MLL4	BRCA1	BRCA2
TCGA-09-0364	G	G	G	3G	L	G
TCGA-09-0365	G,S	G	G	3G,1S	-	-
TCGA-09-0366	S	G	G	2G,1S	-	-
TCGA-09-0367	G	G	G	3G	-	-
TCGA-09-0369	S	G	G	2G,1S	-	-
TCGA-13-0714	G,L,S	G	-	2G,1L,1S	-	-
TCGA-13-0717	G,L,S	G	G	3G,1L,1S	-	-
TCGA-13-0723	L	G	G	2G,1L	-	-
TCGA-13-0724	S	G	G	2G,1S	-	-
TCGA-13-0725	G	L	G	2G,1L	-	-
TCGA-13-0727	G,L,S	G	G	3G,1L,1S	-	-
TCGA-13-0730	G	G	G	3G	S	-
TCGA-13-0751	G,S	G	G	3G,1S	-	-
TCGA-13-0755	G,S	G	G	3G,1S	-	-
TCGA-13-0757	G	G	G	3G	-	-
TCGA-13-0758	G,L,S	G	G,S	3G,1L,2S	-	-
TCGA-13-0760	G,S	G	-	2G,1S	-	-
TCGA-13-0761	G	G	-	2G	S	G,L
TCGA-13-0762	S	G	G	2G,1S	G,L	-
TCGA-13-0765	G,L	G	G	3G,1L	-	-
TCGA-13-0766	G,S	G	-	2G,1S	-	G
TCGA-24-1562	S	-	-	1S	-	G
	72.7% germline	90.9% germline	77.3% germline	95.5% germline	4.5% germline	18.2% germline

High co-occurrences and percentage of germline mutations of CHEK2, RPS6KA2 and MLL4 suggest that these genes could be potential markers for inherited risk of disease development.

A*STAR BII

Key genes involved in etiology of various ovarian cancer subtypes

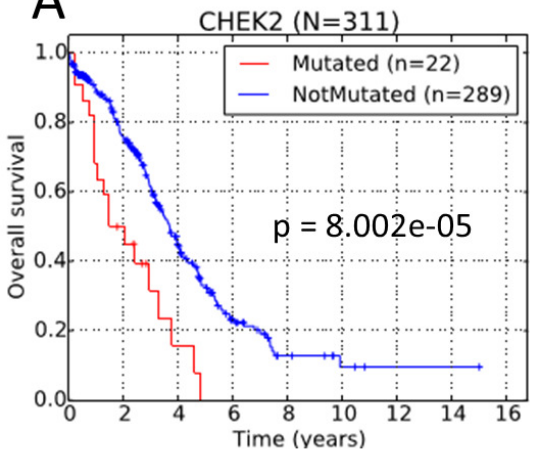


Our proposed model (conditional on the presence of TP53 mutations):

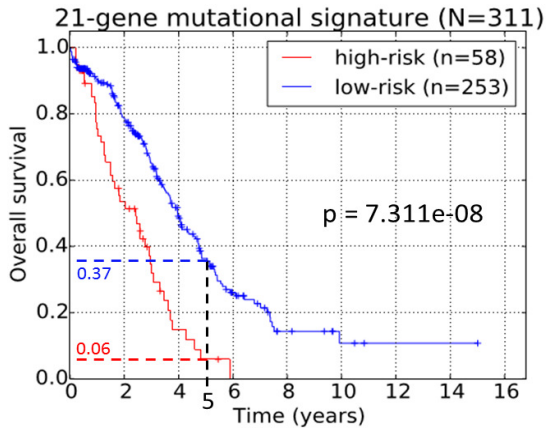
→ 7% of HG-SOC could be due to inherited factors associated with germline CHEK2 mutations

→ 11.6% of HG-SOC could be due to somatic mutations of our signature genes.

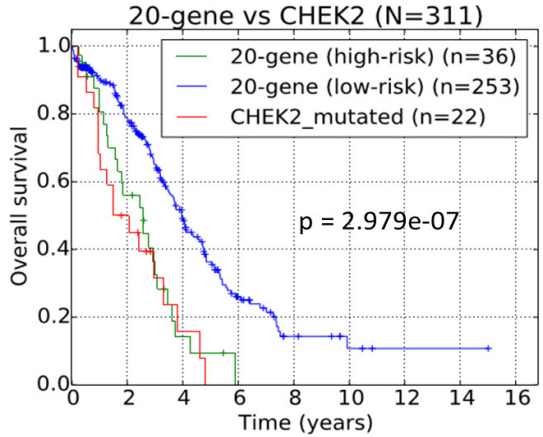
A



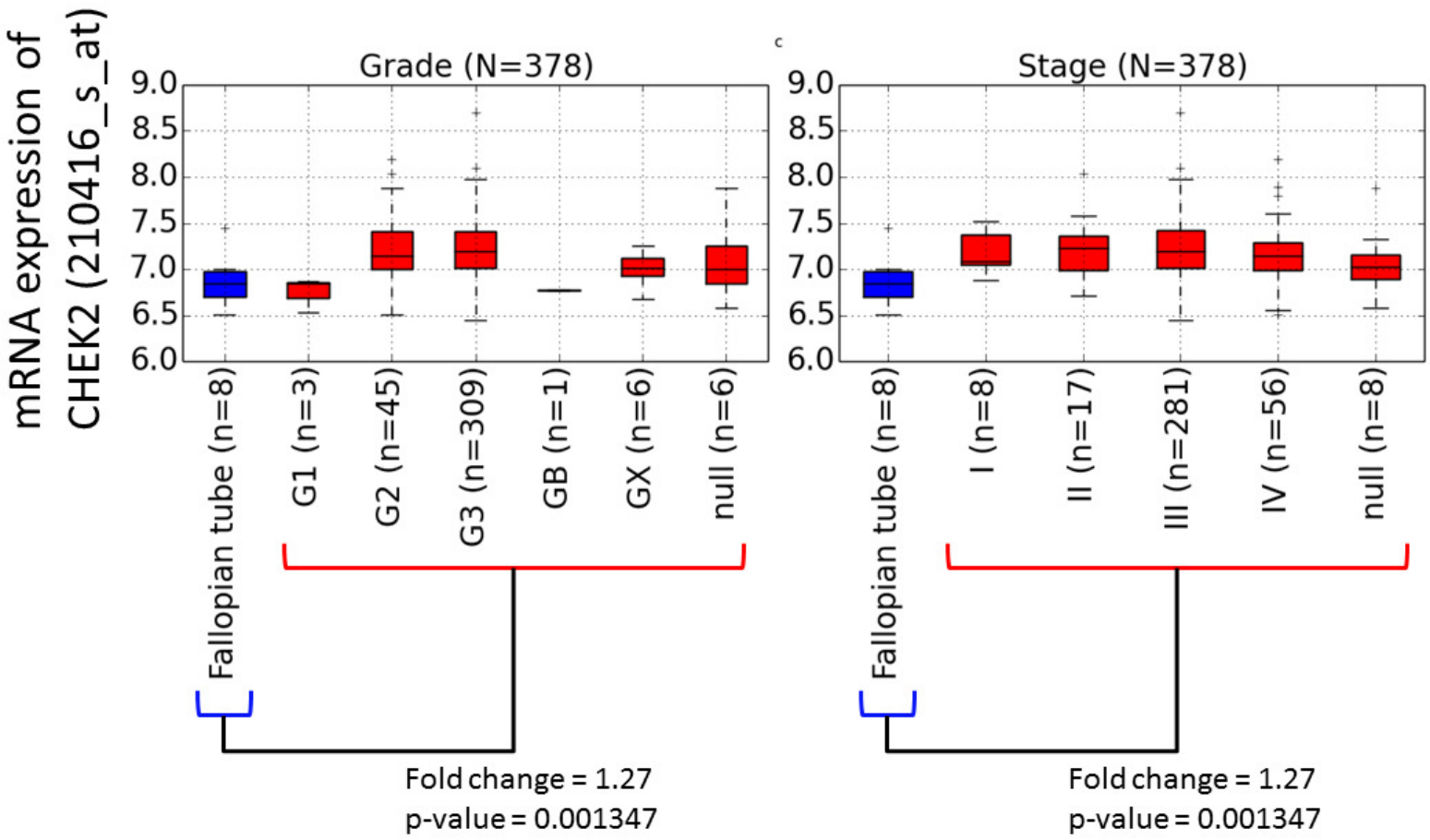
B



C



Expression level of CHEK2 could be potential diagnostic biomarker of HG-SOC relative to normal fallopian tube tissues

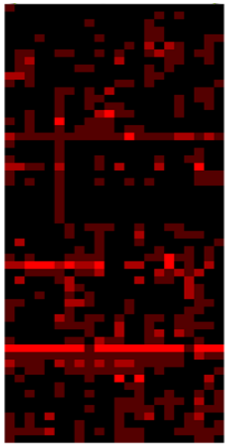


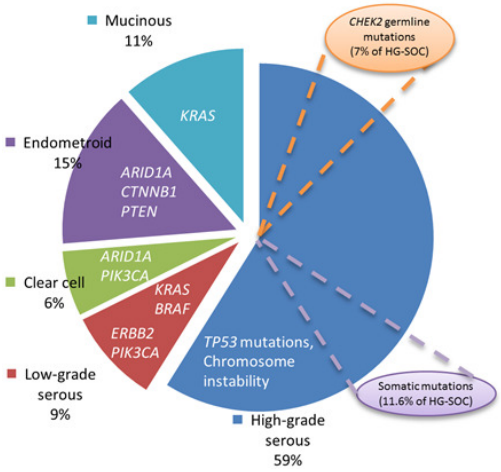
CHEK2 mRNA is up-regulated in HG-SOC samples with reference to normal fallopian tube tissues, suggesting potential of the use of CHEK2 mRNA expression as part of HG-SOC diagnostics.

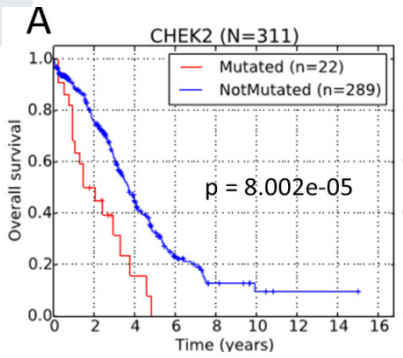
A*STAR BII

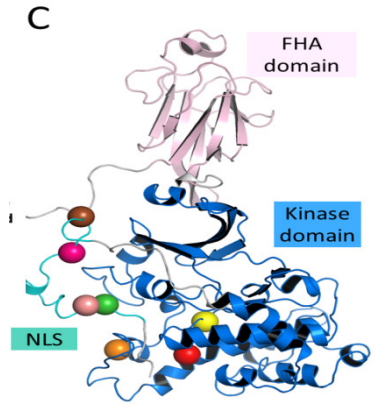
Summary

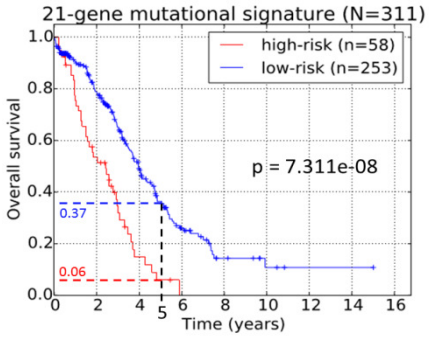
- We identified two novel sub-classes of HG-SOC which are characterized via either germline mutation cluster representing by ***CHEK2***, ***RPS6KA2*** and ***MLL4*** or mostly somatic mutation gene cluster of the other relatively-rare mutated genes.
- Mutations in the genes of these gene sets are related to very poor personalized prognosis, drug resistance and in total affect ~19% HG-SOC patients.
- Mutation of CHEK2 nuclear localization signal could affect nuclear translocation of the mutated protein and may lead to the dysfunctions in DNA repair, cell cycle and apoptosis pathways.
- Expression level of CHEK2 mRNA might be a positive biomarker of early detection of HG-SOC.
- Our findings suggest a development of new diagnostic test(s) for (i) early diagnostics of the women with inherited risk of ovarian cancer, (ii) accurate identification of specific rare mutations associated with poor prognosis and drug resistance, (iii) personalized therapeutic strategies reducing hereditary OC risk.

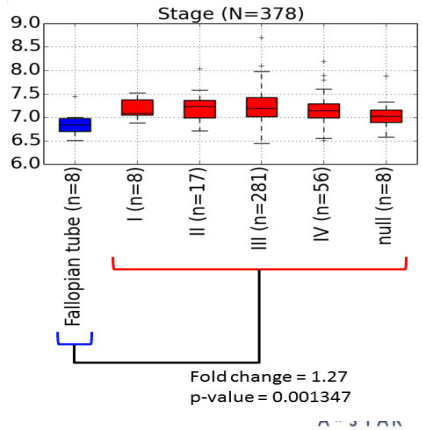












Latest news!

A talent cancer-sniffing dog, Tsunami (University Pennsylvania, USA), detected ovarian cancer by smelling urine samples with 90+% accuracy. This study has being funded by an $80,000 grant from the Madison, N.J.-based Kaleidoscope of Hope Foundation.

Acknowledgement

- Ow Ghim Siong
- Gloria Fuentes
- Ivshina Anna


OW Ghim Siong


Gloria FUENTES


Anna IVSHINA

- Chandra Verma
- Ted Hupp
- Tang Zhiqun (Grace)
- Arsen Batagov
- Yenamandra Syria Pavan
- Frank Eisenhaber

A*STAR BII

Thanks for attention!