N.I. Pirogov Russian National Research Medical University
Department of Molecular Biology and Medical Biotechnology, RNRMU, Moscow, Russia
Fox Chase Cancer Center, Philadelphia, USA

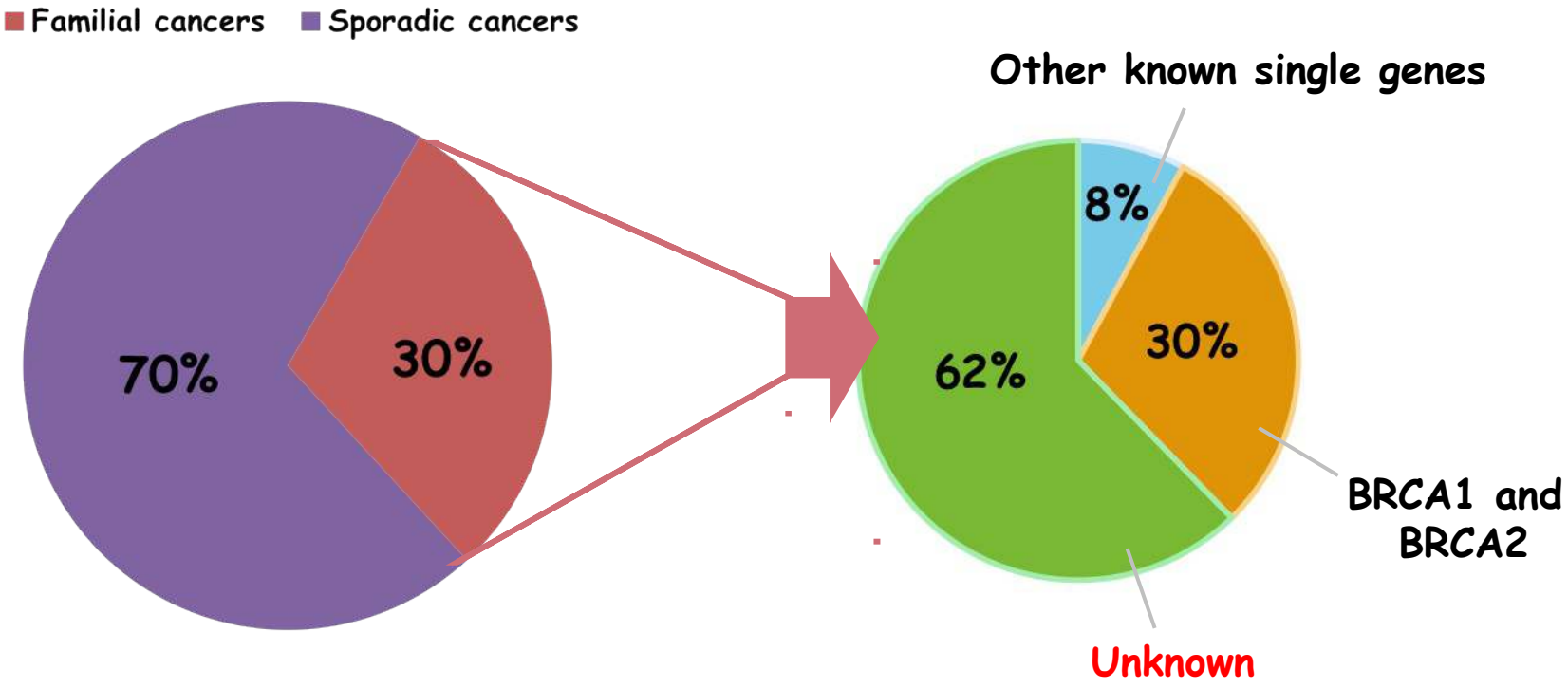
Characterization of *C6orf97*, a Novel Breast Cancer Susceptibility Candidate Gene

Natalia Baulina

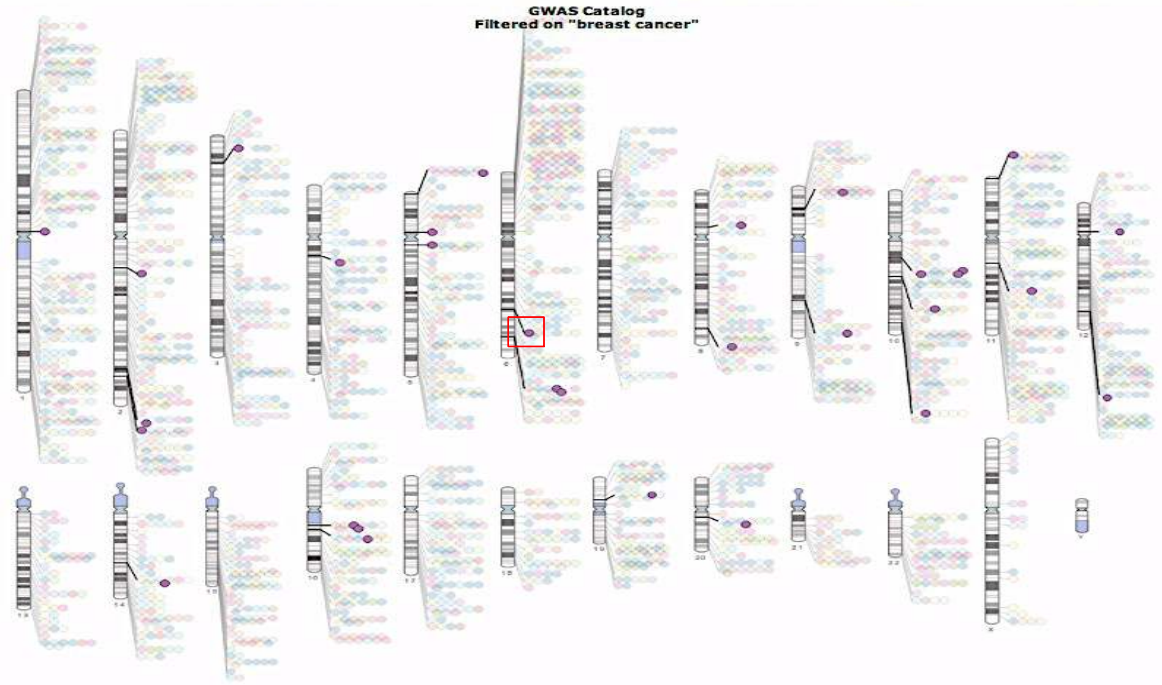
Research advisors: Olga Kulakova, PhD, associate professor
Xiaowei Chen, PhD, associate professor
Head of department: Olga Favorova, professor

Breast Cancer (BCa):

- **BCa** is a disease in which malignant (cancer) cells form in the tissues of the breast.
- The most common cancer and the second most common cause of cancer related death in women
- **Risk factors:** gender, aging, **genetic risk factors,** family history of BCa, race and ethnicity, menstrual periods etc.



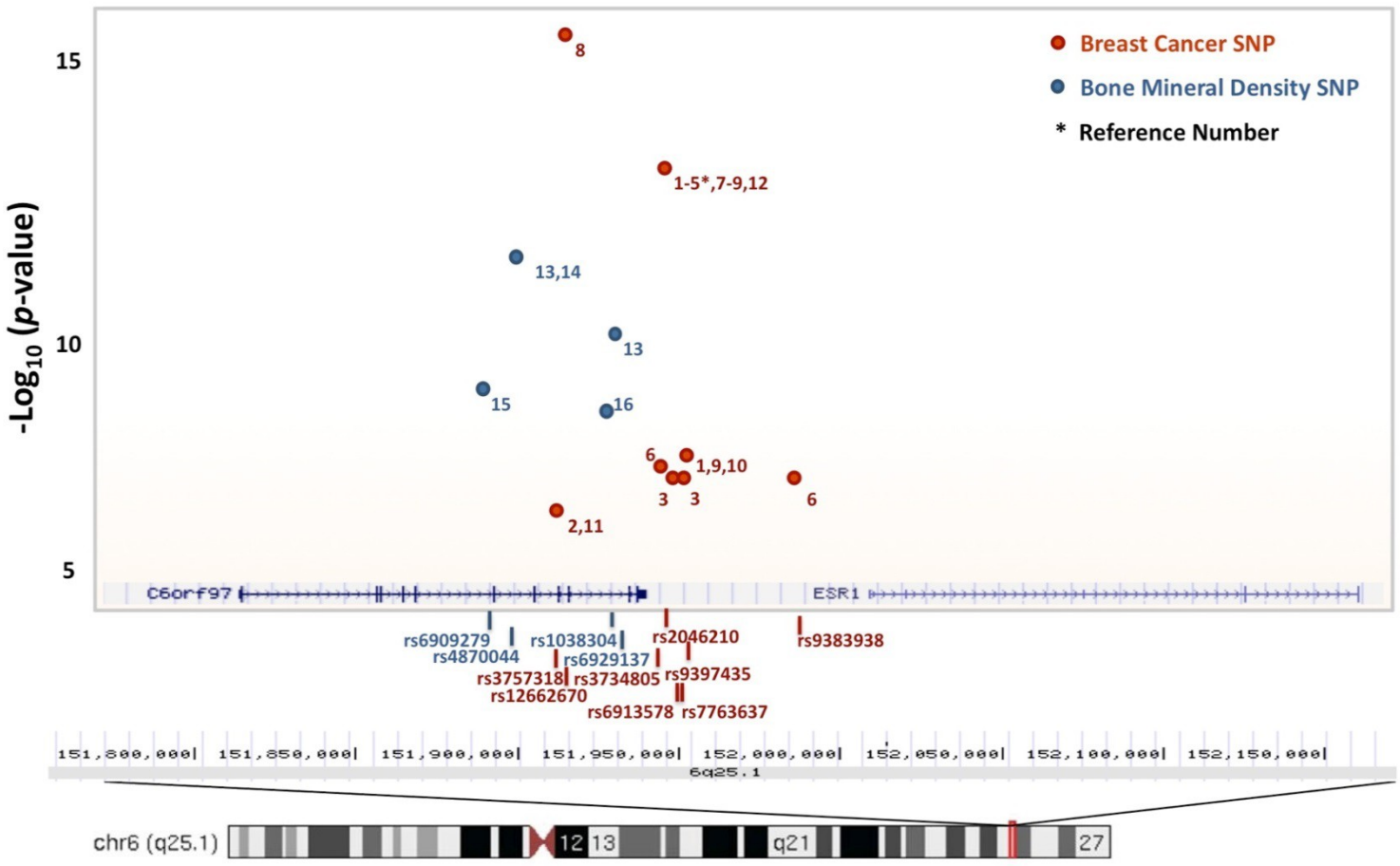
Genome-Wide Association Studies (GWAS) breast cancer susceptibility loci



This diagram shows all 32 SNPs, associated with BCa with p-value ≤ 5.0×10^{-8} (red dots), published in the GWAS catalogue to the end of May 2014.

- **Several GWAS-identified single nucleotide polymorphisms (SNPs) are located on the 6q25.1 chromosome. This locus contains *C6orf97 gene* (Chromosome 6 open reading frame 97), which encodes CCDC170 protein (coiled-coiled domain containing protein 170)**
- ***C6orf97* showed allele-specific expression in human BCa cells**

SNPs associated with BCa at *C6orf97-ESR1* region by *GWAS*



C6orf97-ESR1 GWAS References

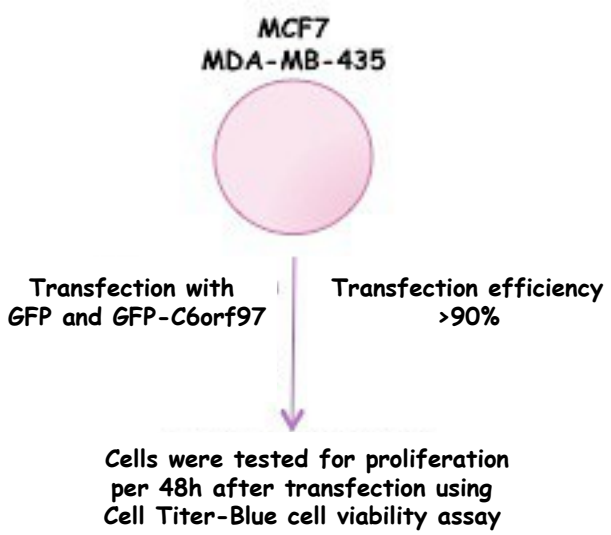
1. Antoniou *et al*, 2011
2. Barzan *et al*, 2013
3. Cai Q *et al*, 2011
4. Chan *et al*, 2012
5. Dai *et al*, 2012
6. Fletcher *et al*, 2011
7. Han *et al*, 2011
8. Hein *et al*, 2012
9. Mulligan *et al*, 2011
10. Stacey *et al*, 2010
11. Turnbull *et al*, 2010
12. Zheng *et al*, 2009
13. Ichikawa *et al*, 2010
14. Koller *et al*, 2013
15. Paternoster *et al*, 2013
16. Styrkarsdottir *et al*, 2008

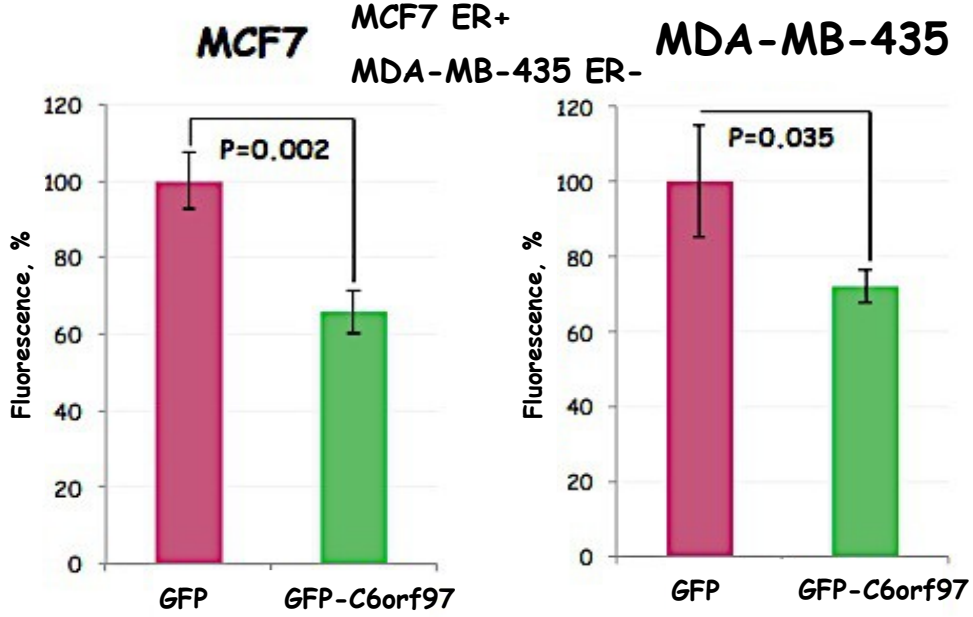
GWAS and related studies suggest that the most BCa and osteoporosis associated polymorphisms of this risk locus are located in *C6orf97* gene

Aim of our study:

Identification of the role of unknown *C6orf97* gene in BCa development

Question 1: What does *C6orf97* gene do in BCa cells ?
Proliferation test



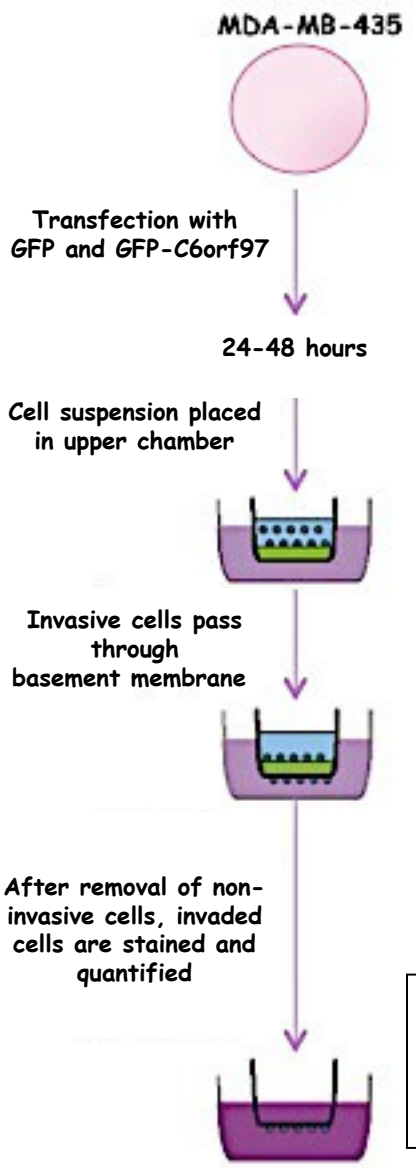


The level of proliferation in MCF7 and MDA-MB-435 BCa cell line, overexpressed with *C6orf97* gene

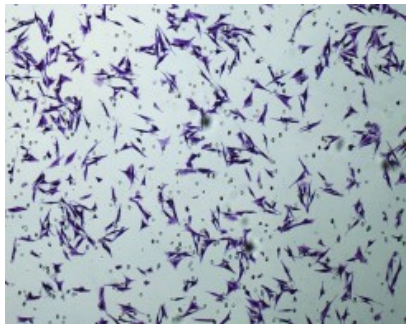
Overexpression of *C6orf97* gene **inhibits cell proliferation** in BCa cell lines independently from ERa production

Question 1: What does *C6orf97* gene do in BCa cells ?

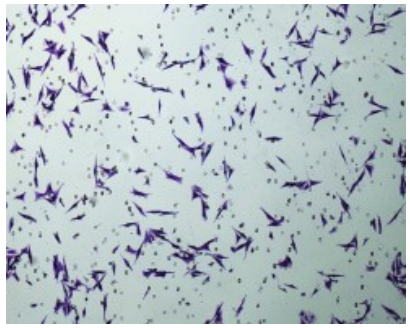
Migration test

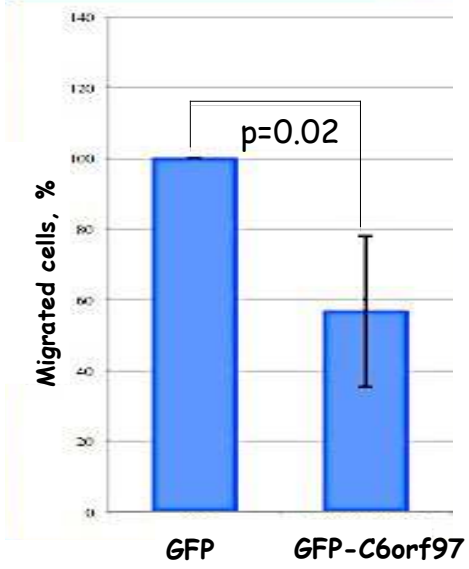


GFP



GFP-C6orf97



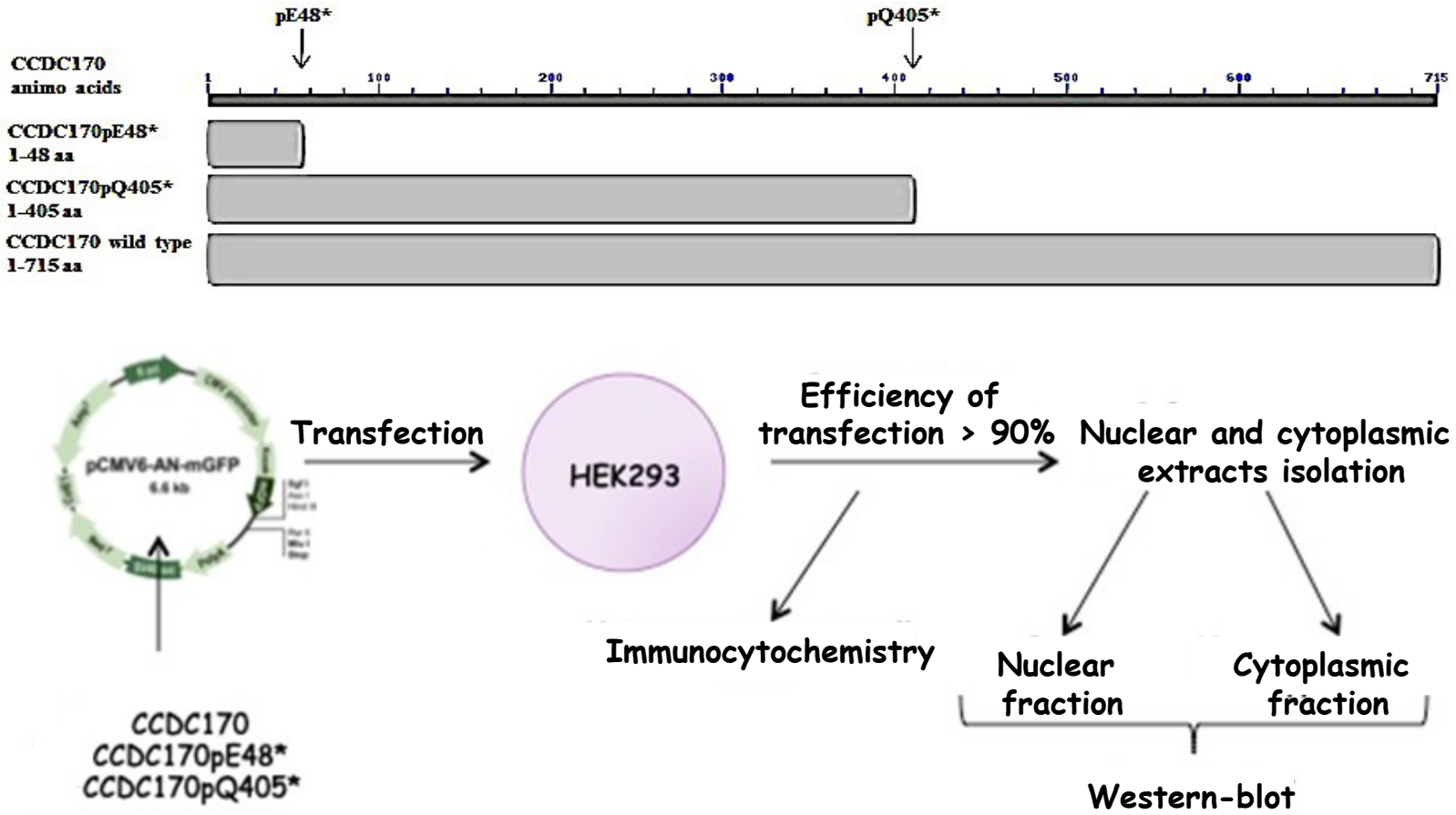


Overexpression of *C6orf97* gene prevents BCa cell migration *in vitro*

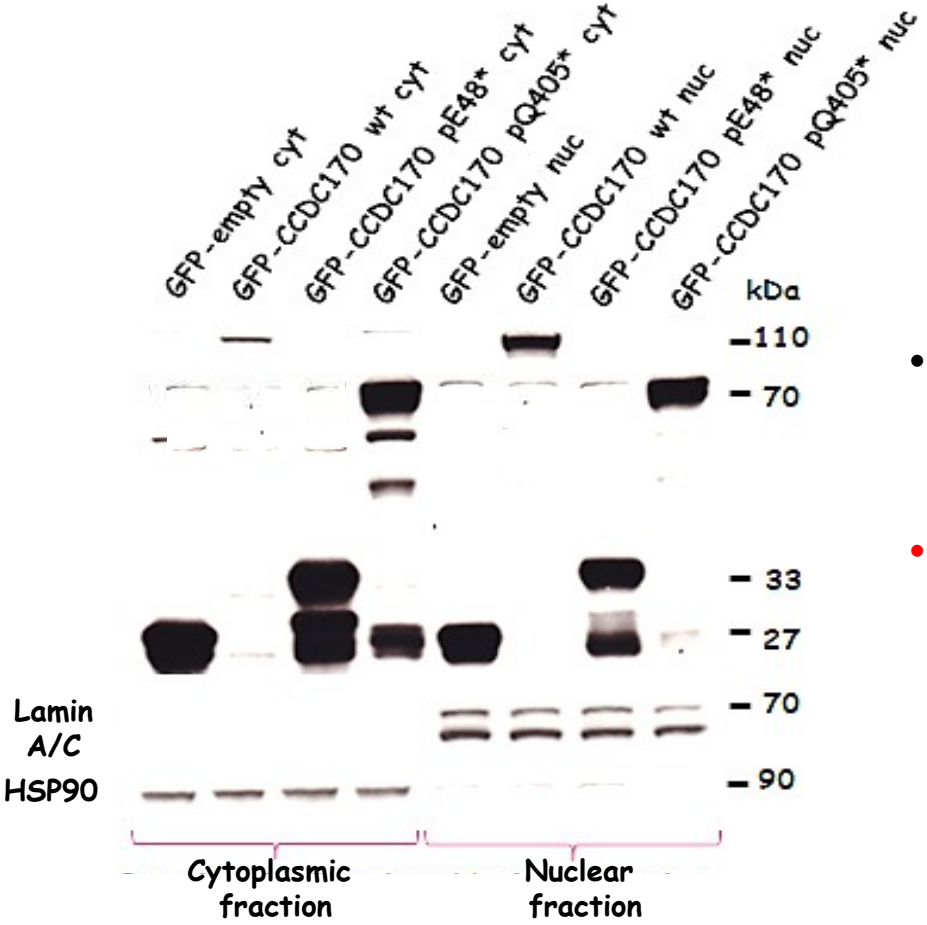
Migration and proliferation tests indicate that *C6orf97* is tumor suppressor gene

Question 2: What is a cellular localization of CCDC170 protein, encoded by *C6orf97* gene?

Schematic images of proteins produced by cells after transfection with constructs, containing *C6orf97* wild type gene or truncated forms of this gene



Western-blot analysis of cytoplasmic and nuclear fractions of HEK293, transfected with investigated constructions

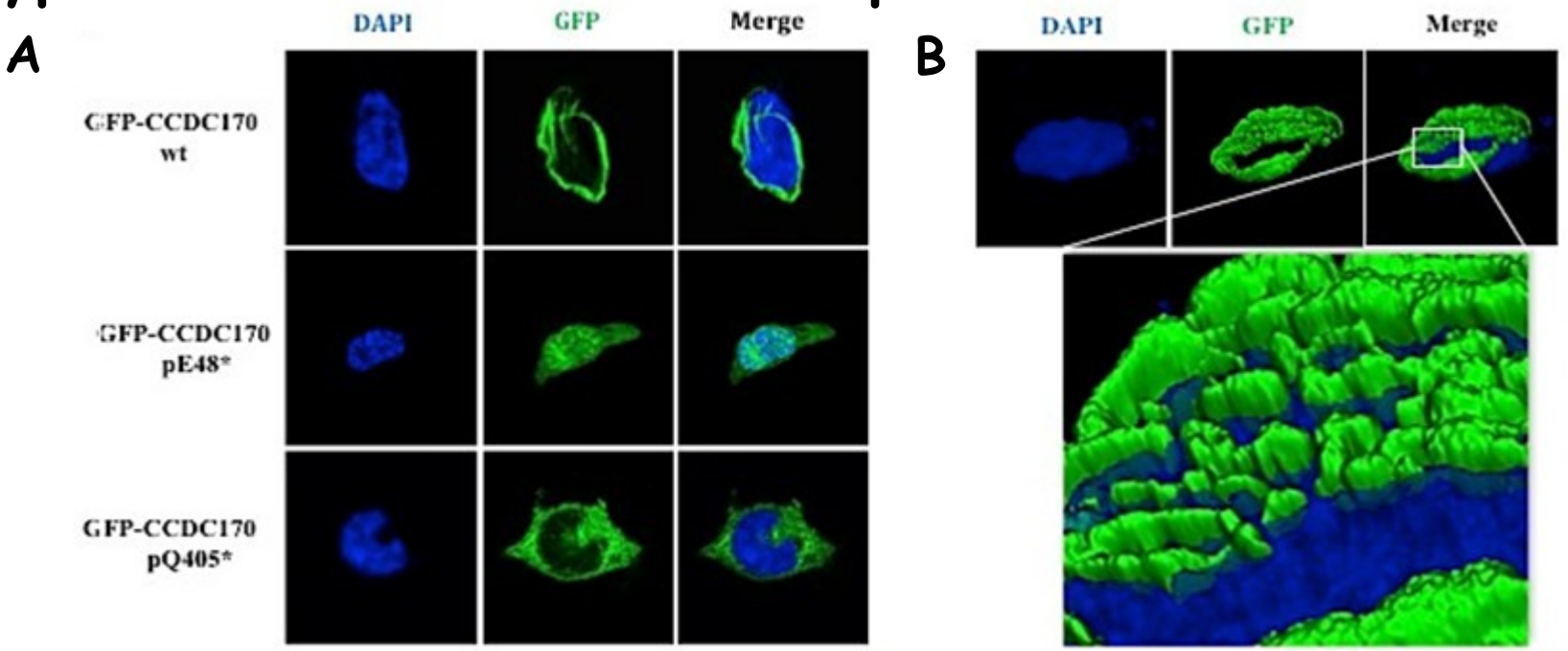


For visualization of CCDC170 wild type protein and truncated forms we used GFP antibody

- Protein CCDC170 wild type is predominantly localized in nuclear fraction;
- Truncated forms are localized in both cytoplasmic and nuclear fractions.

LaminA/C and HSP90 were used as a nuclear or cytoplasmic marker protein, respectively.

Immunocytochemistry: confocal images of CCDC170 wild type and truncated CCDC170 proteins



The signals of CCDC170 wild type and truncated proteins (green) were detected using GFP-tags. Nuclei were counterstained with DAPI (blue).

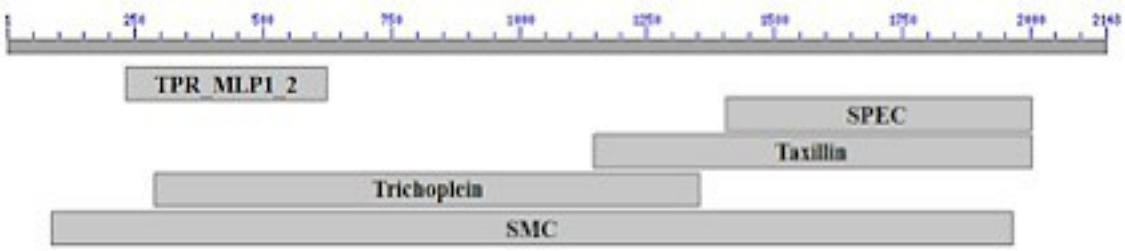
A. Localization of CDC170 wild type and truncated CCDC170 proteins

B. 3D CCDC170 wild type cellular distribution

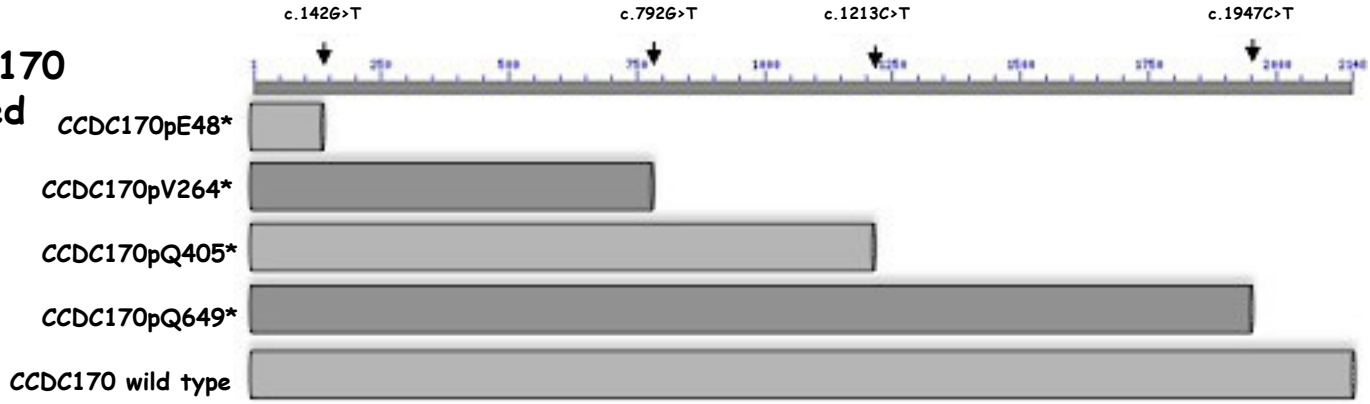
- **CCDC170 wild type is localized to the periphery of the nucleus, more likely in the nuclear envelope.**
- Truncated form **GFP-CCDC170pE48*** is diffusively distributed in cell.
- Truncated form **GFP-CCDC170pQ405*** is localized in cytoplasm and, perhaps, in nuclear envelope.

Question 3: Which terminal end of CCDC170 protein is critical for its cellular localization?

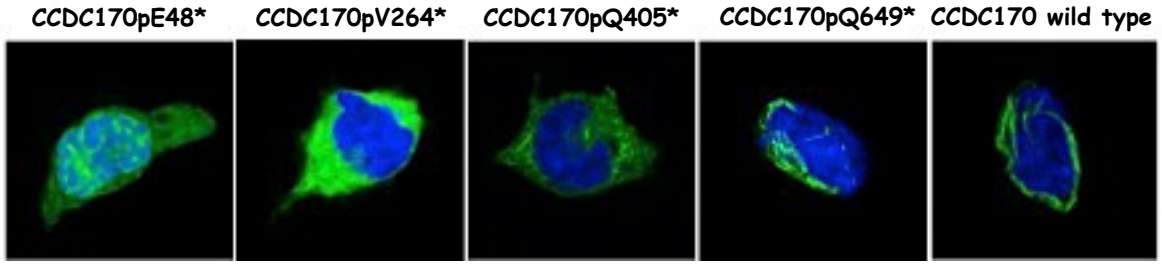
Domain organization of CCDC170



Schematic view of CCDC170 wild type and truncated proteins

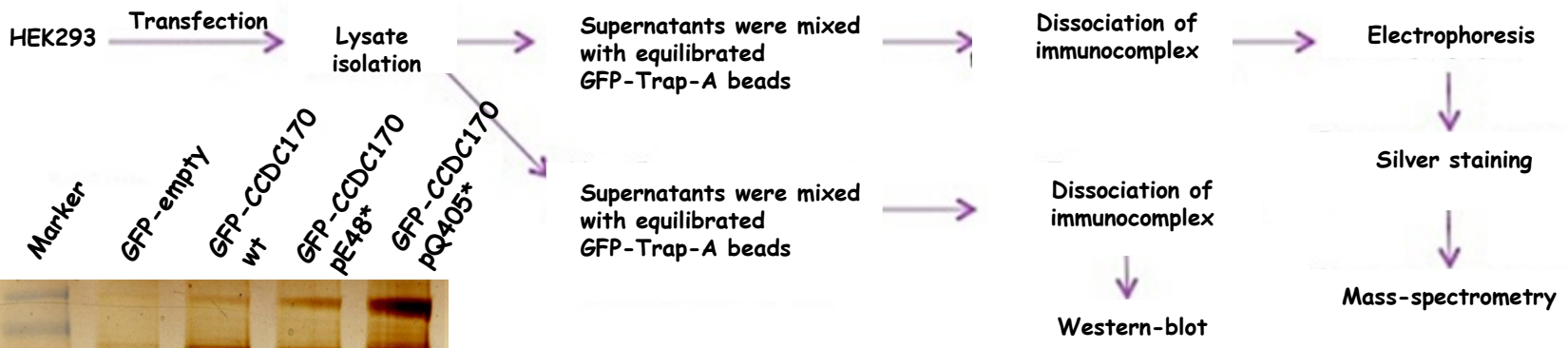


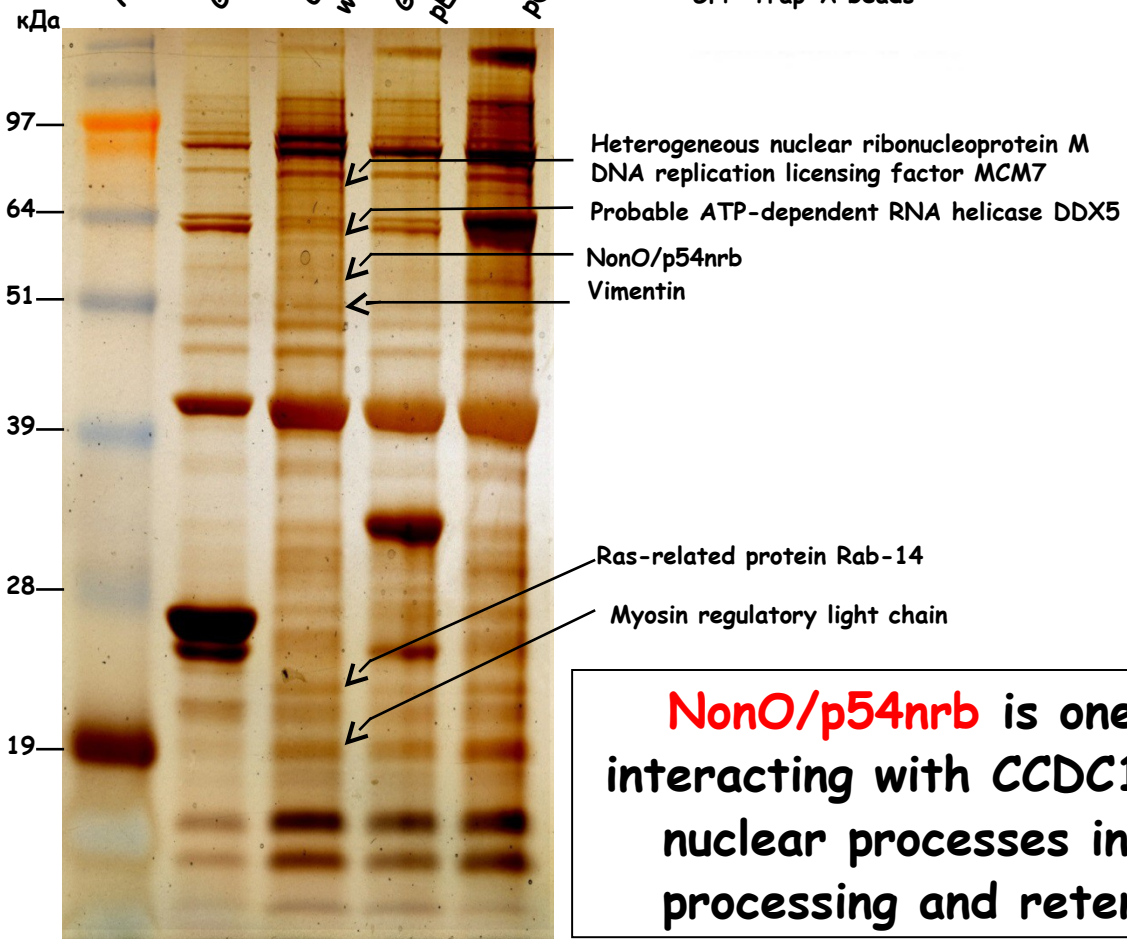
Confocal immunofluorescence images of CCDC170 wild type and truncated forms

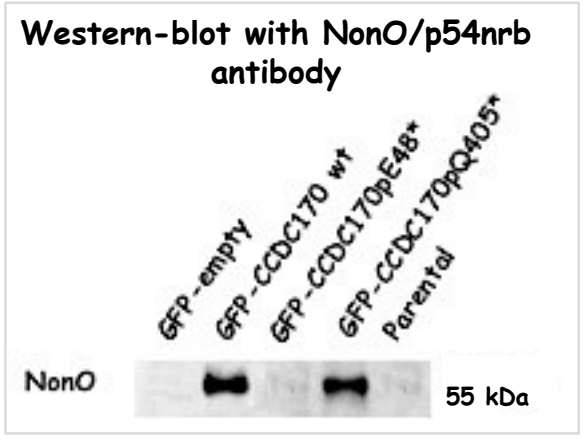


C-terminal end of CCDC170 protein is crucial for cellular localization, presumably in nuclear envelope

Question 4: What are the proteins interacting with CCDC170?







NonO/p54nrb is one of the validated proteins, interacting with CCDC170, which is involved in many nuclear processes including transcription, RNA processing and retention of hyperedited RNAs.

Conclusions:

1. Gene *C6orf97* inhibits BCa cell proliferation and migration,
2. Protein CCDC170, encoded by *C6orf97* gene, is localized to the periphery of the nucleus more probably in nuclear envelope
3. C-terminal end of CCDC170 protein is critical for its localization more probably in nuclear envelope.
4. Protein CCDC170 directly interacts with protein NonO, which is involved in many nuclear processes including transcription, RNA processing, and retention of hyperedited RNAs.

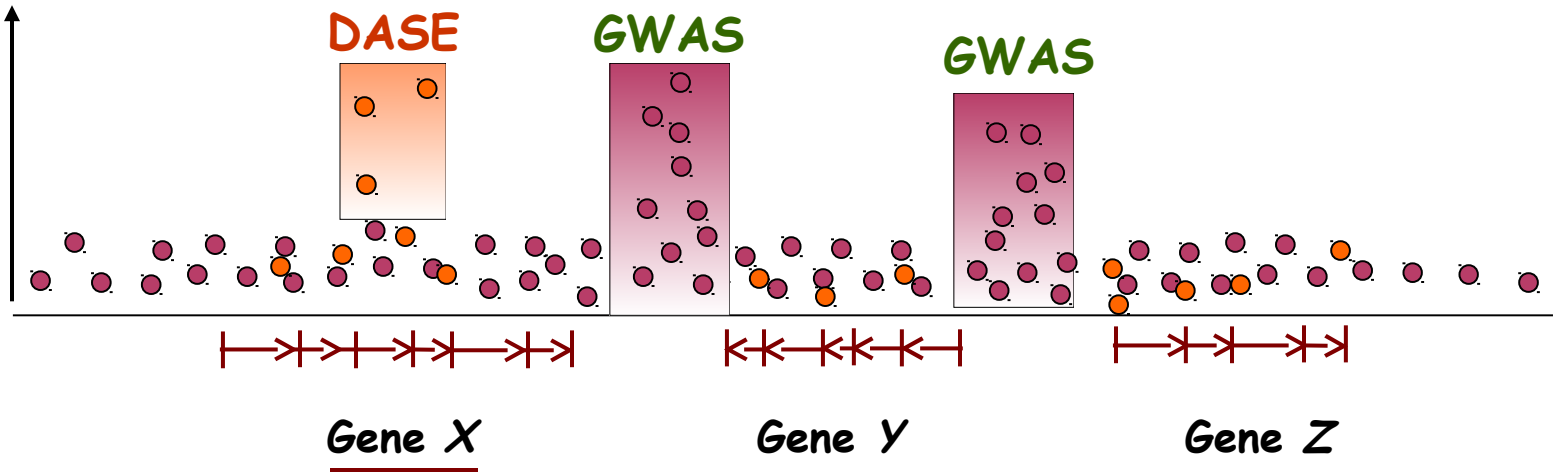
According to our findings, *C6orf97* is a novel breast cancer suppressor gene, involved in post-transcriptional regulation of gene expression

Thank you for your attention!

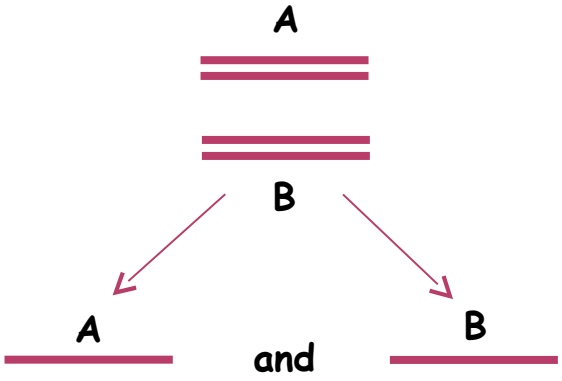
Differential Allele-Specific Expression (DASE)

- DASE is a common phenomenon which has been shown to contribute to phenotypic variability in human and more recently to the pathogenesis of cancer.
- DASE is a result of epigenetic modifications and is associated with X-chromosome inactivation and genomic imprinting and also presents among non-imprinted autosomal genes caused by cis-regulatory variations. As a result the ratio between the expression of two alleles is changed - allelic imbalance.
- **Variation in gene expression is highly heritable and a significant determinant of human disease susceptibility**

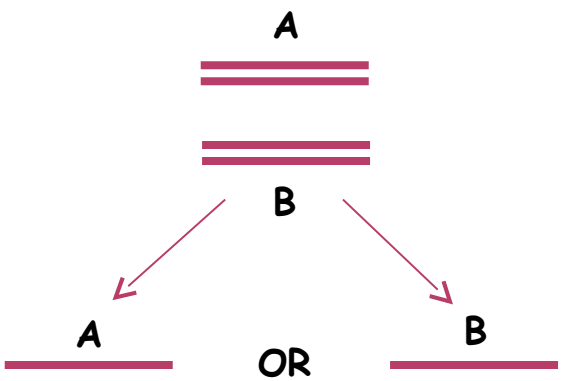
Overlying the resulting DASE maps with GWAS data allowes us to identify novel BC candidate loci



Differential Allele-Specific Expression (DASE):

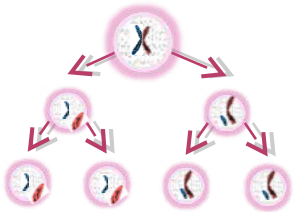


Bi-allelic expression

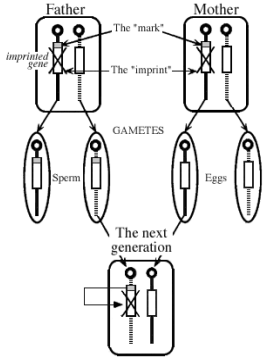


Differential Allele-Specific Expression (DASE):

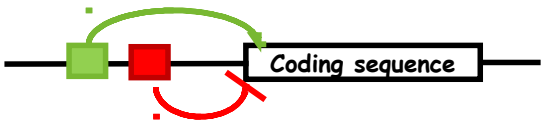
X-chromosomal inactivation



Genomic imprinting

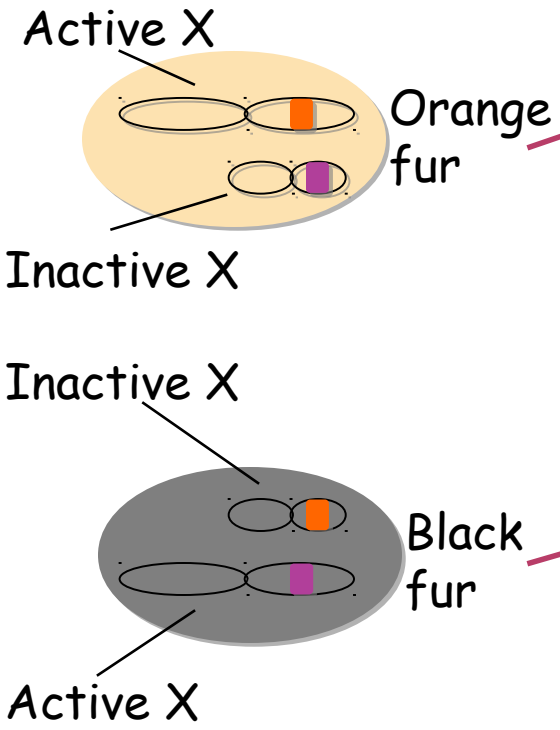


Cis-regulation



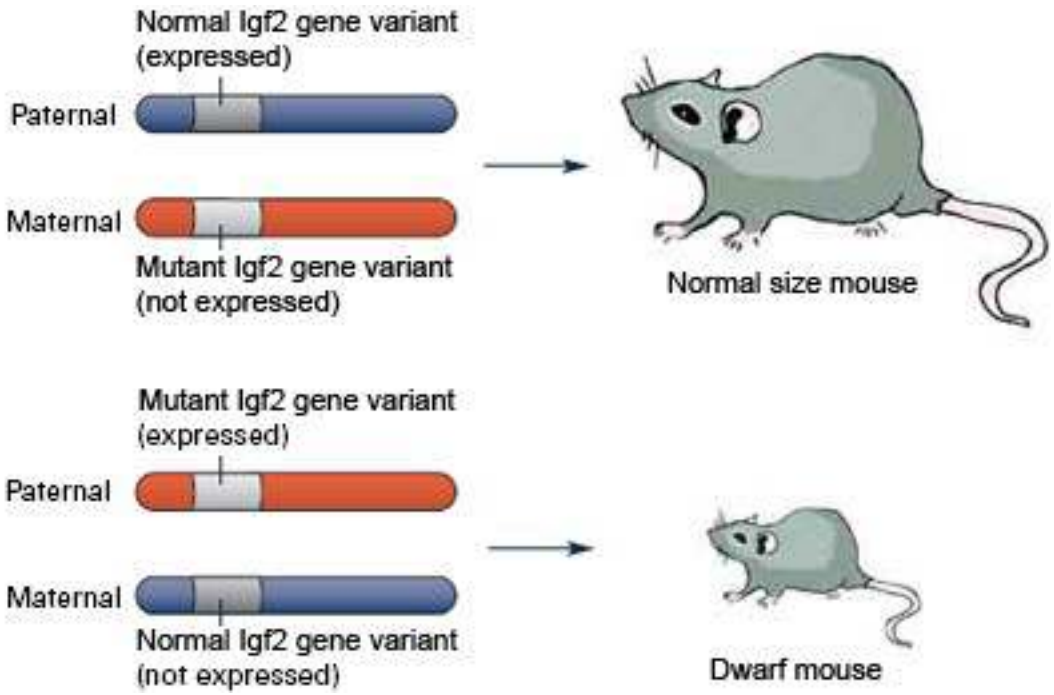
DASE: X-chromosome inactivation

Two cell populations in adult cat:



The patchy distribution of color on tortoiseshell cats results from the random inactivation of one X chromosome in females

DASE: Genomic imprinting



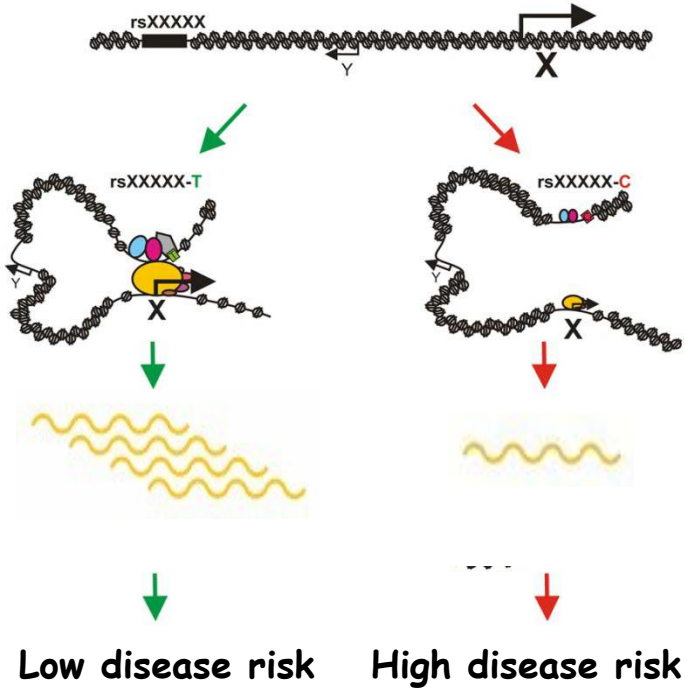
Igf2 is a gene in mice and humans which is expressed only from the paternal allele. It is a potent growth factor, and when disrupted in mice leads to tiny pups.

Image courtesy of the NIH.

DASE: *Cis*-regulation

Table 1:
Examples of regulatory mutations associated with human disease

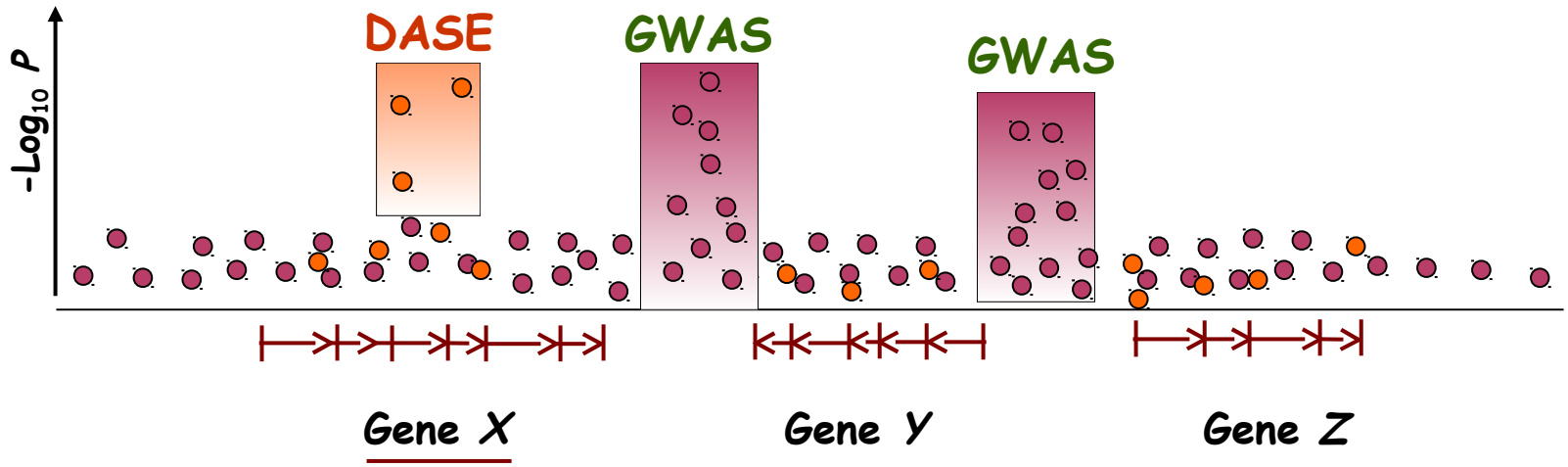
Gene	Disease	Location of rSNP	TF-binding site affected	References
HBB	β-thalassemia	Promoter	Several (TATA, CACCC, EKLF)	[13]
F9	Hemophilia B	Promoter	Several (HNF4, C/EBP)	[13]
LDLR	Familial hypercliolesterolemia	Promoter	Several (SP1, SRE repeat)	[13]
CollAI	Osteoporosis	Intron 1 (+2kb)	SP1 (gain)	[14]
RET	Hirschprung	IntronI (+9.7 kb)	Unknown	[23]
HBA	α-thalassemia	Upstream (−13 kb)	GATA1 (gain)	[25]
SHH	Preaxial polydactyly	Upstream	Unknown	[29]



Douglas J. Epstein, 2009

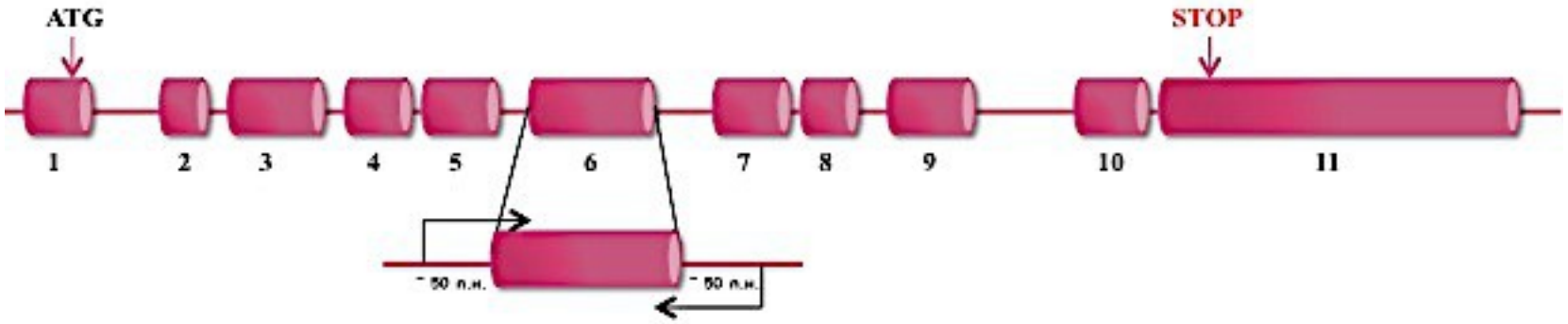
Variation in gene expression is highly heritable and a significant determinant of human disease susceptibility

Global DASE patterns in mammary transcriptome: Maximizing the Yield of True Positive Loci from BCa GWAS



C6orf97 exon mutation screening

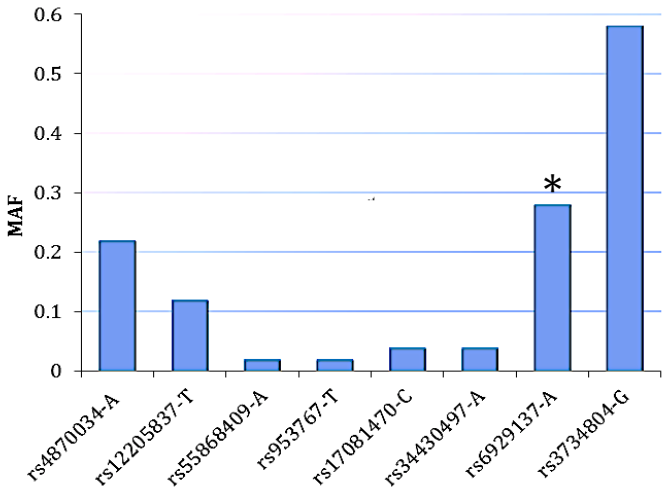
- Amplification of 11 exons of *C6orf97* gene on 25 non-Jewish BRCA1 and BRCA2 negative affected probands using PCR



- Sequencing of amplified fragments

Exon	SNP	DNA sequence changes	Protein sequence changes
3	rs4870034	c.321A>G	p.Glu107=
6	rs12205837 rs55868409 rs953767	c.806C>T c.1033G>A c.971T>C	p.Ala269Val p.Glu345Lys p.Phe324Ser
8	rs17081470	c.1317T>C	p.Leu439=
9	rs34430497	c.1658G>A	p.Arg553Gln
10	rs6929137	c.1810G>A	p.Val604Ile
11	rs3734804	c.2047G>A	p.Val683Ile

SNPs, identified from mutation screening



*** - SNP, associated with osteoporosis**