Biologists are from Venus,

Mathematicians are from Mars,

They cosegregate on Earth,

And conditionally associate to create a DIGGIT.

Identification of Causal Genetic Drivers of Human Disease through Systems-Level Analysis of Regulatory Networks

JAMES C. CHEN
MARIANO J. ALVAREZ
FLAMINIA TALOS
HARSHIL DHRUV

Motivation

1. Identification of Driver Mutations is usually performed with **statistical models.**
2. These models can identify only the **highly penetrant and frequent driver events.**
   - To achieve **statistical power** (in context of multiple hypothesis-testing correction), these models need large cohorts and/or large effect sizes.
3. Moreover, these models typically do not provide **mechanistic insight.**
4. On the other hand, Gene-based biochemical studies can **provide insight into regulatory mechanisms** but **do not scale**.

Problem

Can we identify genetic determinants of a disease:

- ✓ Can we go beyond the highly penetrant and frequent driver events
- ✓ While remaining statistically rigorous
- ✓ Without using extremely large cohorts

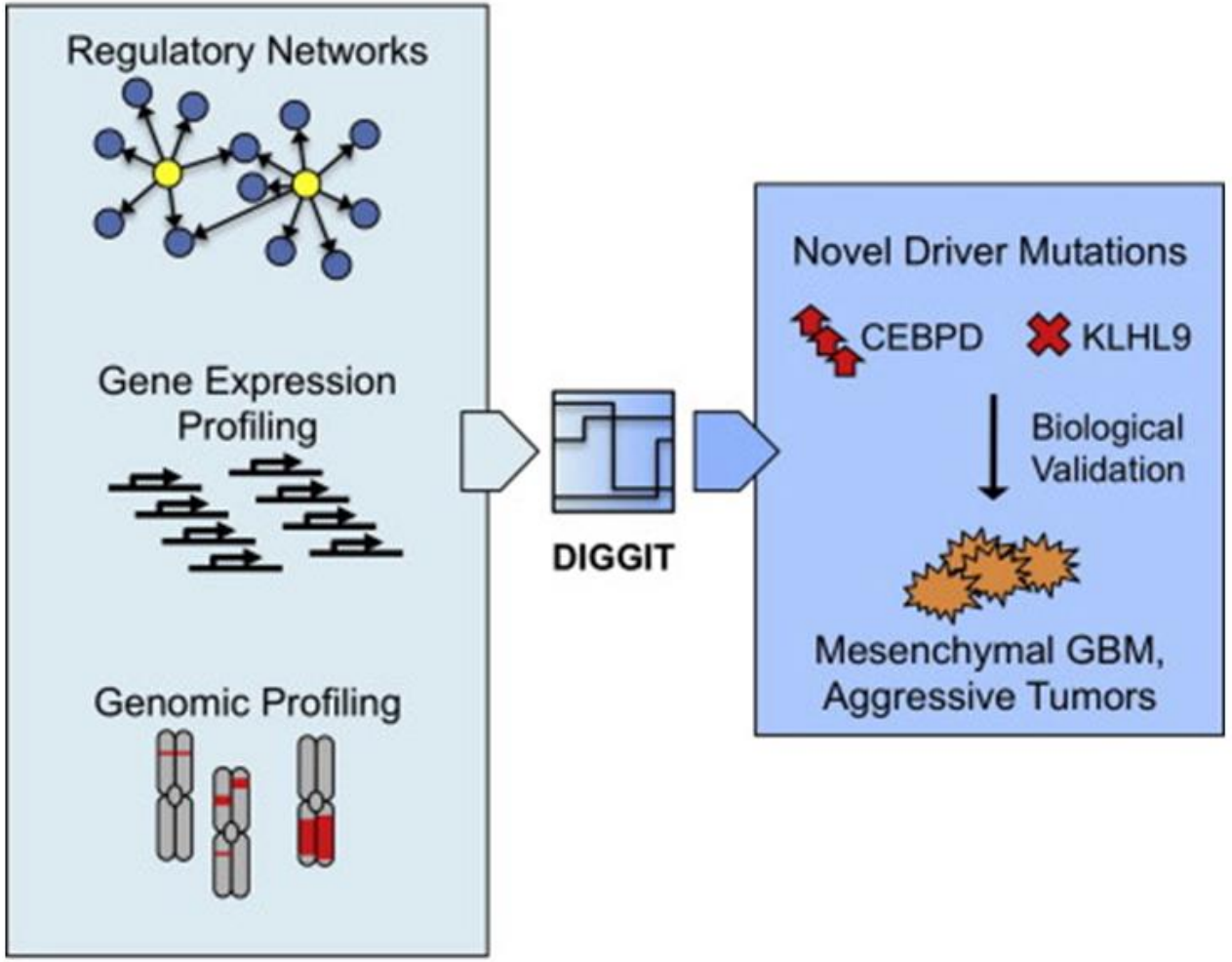
Can such an algorithm provide mechanistic insight into the process by which these genetic determinants play out their effect?

Idea

1. **Overall Idea:**
   - Diverse alteration patterns induce **common aberrant signals**.
   - These signals converge on regulatory modules and associated MR proteins that represent **key regulatory bottlenecks**.
   - **Dysregulation** of these bottlenecks is both necessary and sufficient **for disease initiation/progression**.
   - Once MR proteins and modules representing regulatory bottlenecks are identified, driver **genetic events must be harbored either by these MRs or by their upstream pathways**.
2. Algorithm can identify these **driver genetic events** by systematically exploring regulatory/signaling networks upstream of these MR genes:
   - Approach is likely to **collapse the number of testable hypotheses.**
   - Approach may **provide regulatory clues** to help elucidate associated mechanisms.

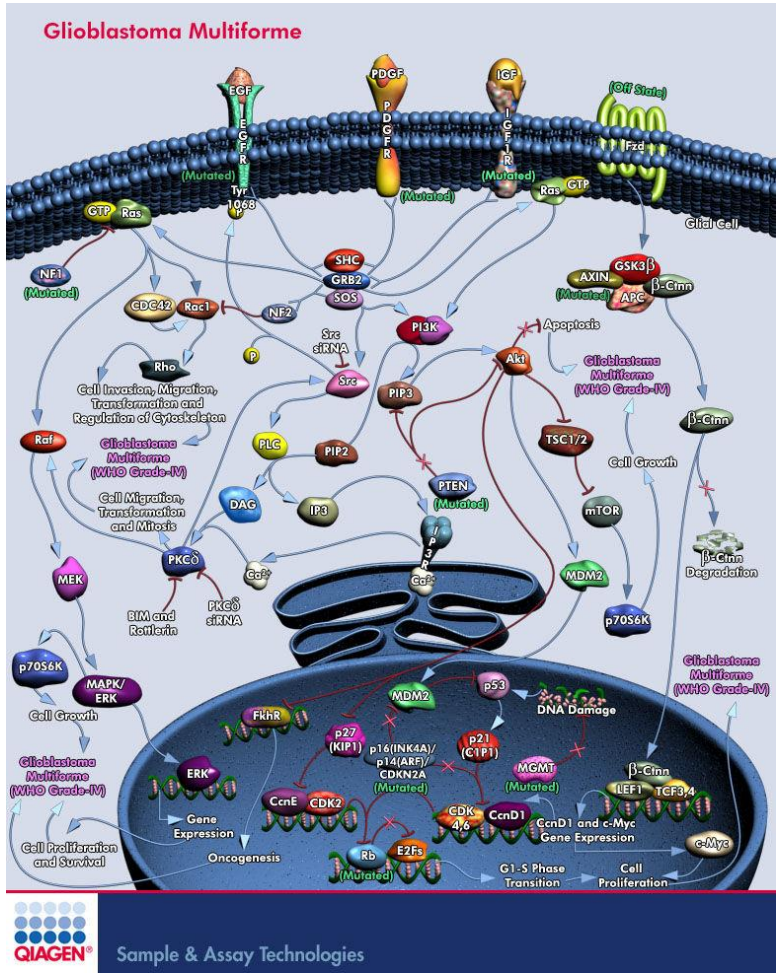
Solution: DIGGIT: Driver Gene Inference by Genetical-Genomics and Mutual Information

DIGGIT: Summary of Findings



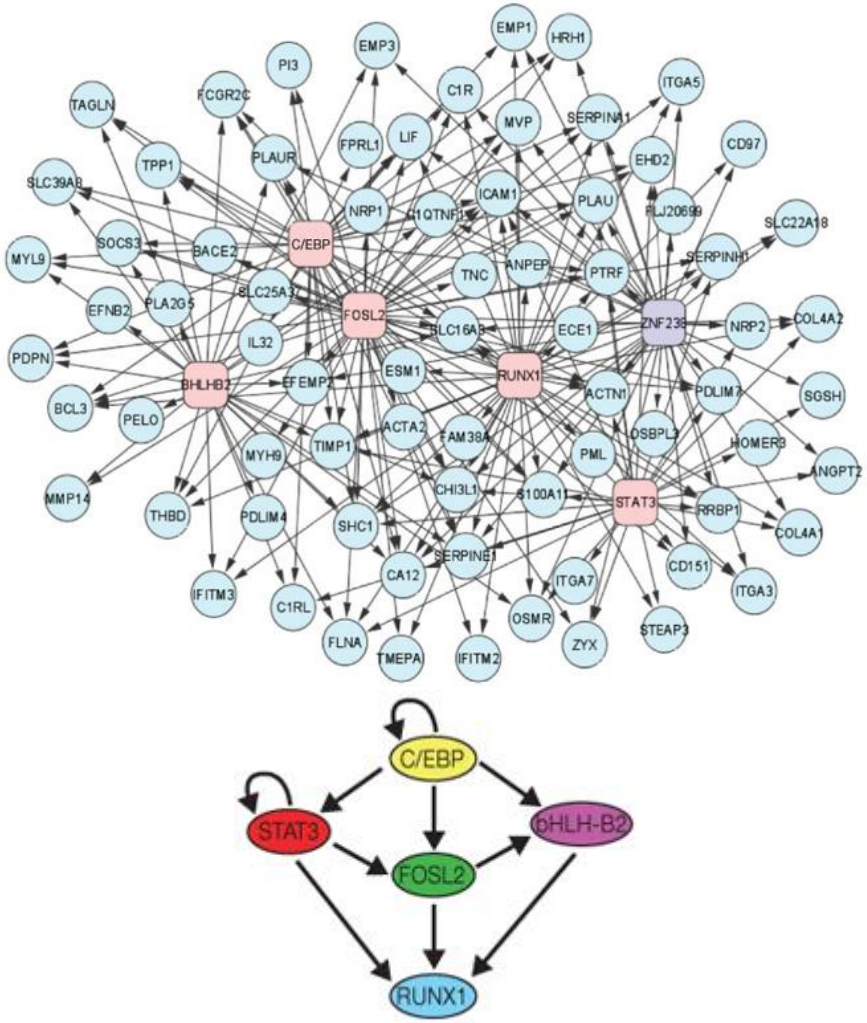
1. Combining cellular networks, gene expression, and genomic data (DIGGIT) **finds novel driver mutations**.
2. Uncovered **KLHL9 deletions** as upstream activators of two previously established Master Regulators of the subtype, C/EBPβ and C/EBPδ.
3. KLHL9 deletions **predict mesenchymal transformation and poorest prognosis** in GBM.
4. KLHL9 post-translationally **regulates CEBPβ/δ**.
5. Rescue of KLHL9 expression **inhibits tumor growth** by inducing degradation of C/EBP proteins and abrogating the mesenchymal signature.
6. DIGGIT can be **used on any genetic disease** with matched expression and genomic data.

MES-GBM: An Ideal Candidate



1. Glioblastoma Multiforme (GBM) is the most common human brain malignancy.
2. Virtually incurable, very aggressive and deadly - average survival of 12–18 months post-diagnosis.
3. Three subtypes associated with expression of mesenchymal, proliferative, and proneural (PN) genes.
4. MES-GBM has the worst prognosis.
5. Despite multiple studies, genetic determinants of MES-GBM are largely elusive.
6. Provides an ideal context to test this rationale, as its established genetic determinants account for < 25% of the patients.

Link to Prior Work



1. In 2010 (*The Transcriptional Network for Mesenchymal Transformation of Brain Tumours)*, reported that aberrant co-activation of the transcription factors (TFs) ***C/EBPβ*, *C/EBPδ*, and *STAT3*** is necessary and sufficient to induce mesenchymal reprogramming in GBM.
2. This suggested that this TF module represents an **obligate pathway** or **regulatory bottleneck** between driver alterations and aberrant mesenchymal program activity.
3. Hypothesize that the **genetic drivers of MES-GBM** are either among these genes or in their upstream pathways.

Mutual Information

Slides borrowed from

University of Wisconsin, Madison (CS 760)

University of Illinois, Chicago (ECE 534)

Entropy

Information theory background

- suppose there are only four types of bikes
- we could use the following code

type	code
Trek	11
Specialized	10
Cervelo	01
Serrota	00

- expected number of bits we have to communicate: 2 bits/bike

Information theory background

- we can do better if the bike types aren't equiprobable
- optimal code uses $-\log_2 P(y)$ bits for event with probability $P(y)$

Type/probability	# bits	code
$P(\text{Trek}) = 0.5$	1	1
$P(\text{Specialized}) = 0.25$	2	01
$P(\text{Cervelo}) = 0.125$	3	001
$P(\text{Serrota}) = 0.125$	3	000

- expected number of bits we have to communicate: 1.75 bits/bike

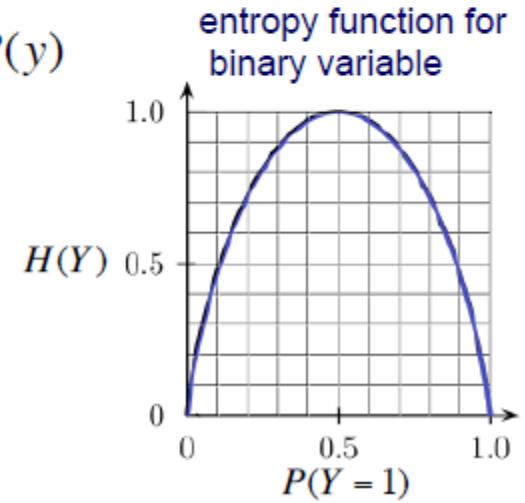
$$-\sum_{y \in \text{values}(Y)} P(y)\log_2 P(y)$$

Entropy

Entropy

- entropy is a measure of uncertainty associated with a random variable
- defined as the expected number of bits required to communicate the value of the variable

$$H(Y) = -\sum_{y \in \text{values}(Y)} P(y) \log_2 P(y)$$



Conditional entropy

- What's the entropy of Y if we condition on some other variable X?

$$H(Y \mid X) = \sum_{x \in \text{values}(X)} P(X = x)\, H(Y \mid X = x)$$

where

$$H(Y \mid X = x) = -\sum_{y \in \text{values}(Y)} P(Y = y \mid X = x) \log_2 P(Y = y \mid X = x)$$

Entropy: Example

The Entropy of a randomly selected letter in an English document is about 4.11 bits. Assuming its probability is as given in the table, we obtain this number by averaging log $1/p_i$ (shown in the fourth column) under the probability distribution (third column)

i	a_i	p_i	$h(p_i)$
1	a	.0575	4.1
2	b	.0128	6.3
3	c	.0263	5.2
4	d	.0285	5.1
5	e	.0913	3.5
6	f	.0173	5.9
7	g	.0133	6.2
8	h	.0313	5.0
9	i	.0599	4.1
10	j	.0006	10.7
11	k	.0084	6.9
12	l	.0335	4.9
13	m	.0235	5.4
14	n	.0596	4.1
15	o	.0689	3.9
16	p	.0192	5.7
17	q	.0008	10.3
18	r	.0508	4.3
19	s	.0567	4.1
20	t	.0706	3.8
21	u	.0334	4.9
22	v	.0069	7.2
23	w	.0119	6.4
24	x	.0073	7.1
25	y	.0164	5.9
26	z	.0007	10.4
27	-	.1928	2.4
		$\sum_i p_i \log_2 \frac{1}{p_i}$	4.1

Table 2.9. Shannon information contents of the outcomes a–z.

Entropy is Important

(A) entropy is the measure of **average uncertainty** in the random variable

(B) entropy is the **average number of bits** needed to describe the random variable

(C) entropy is a lower bound on the **average length of the shortest description** of the random variable

(D) entropy is measured in bits?

(E) $H(X) = -\sum_x p(x) \log_2(p(x))$

(F) entropy of a deterministic value is 0

Mutual Information

- Entropy H(X) is the uncertainty (`self-information'') of a single random variable
- Conditional entropy H(X|Y) is the entropy of one random variable *conditional upon* knowledge of another.
- The average amount of decrease of the randomness of X by observing Y is the average information that Y gives us about X.

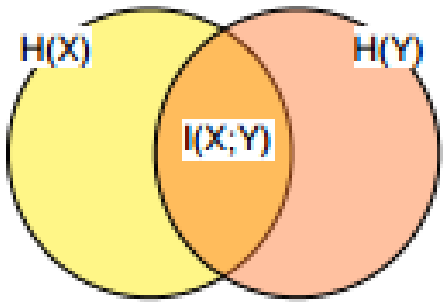
Definition: **The mutual information $I(X;Y)$ between the random variables X and Y is given by**

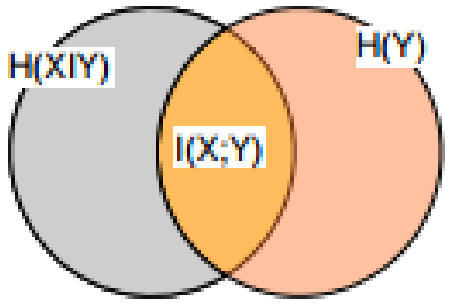
$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \\
&= E_{p(x,y)} \left[\log_2 \frac{p(X,Y)}{p(X)p(Y)} \right]
\end{aligned}
$$

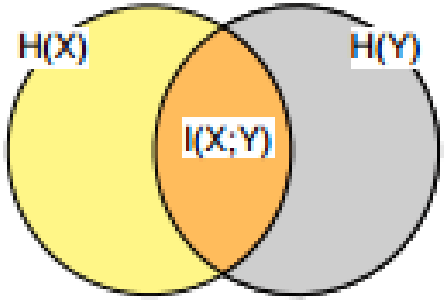
Mutual Information and Entropy

Theorem: Relationship between mutual information and entropy.

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
I(X;Y) &= H(Y) - H(Y|X) \\
I(X;Y) &= H(X) + H(Y) - H(X,Y) \\
I(X;Y) &= I(Y;X) \quad \text{(symmetry)} \\
I(X;X) &= H(X) \quad \text{(“self-information”)}
\end{aligned}
$$

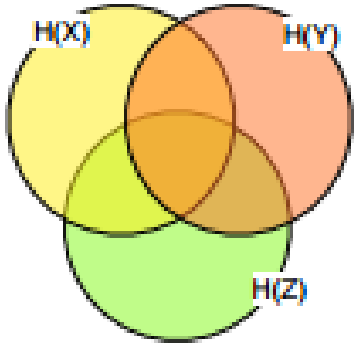






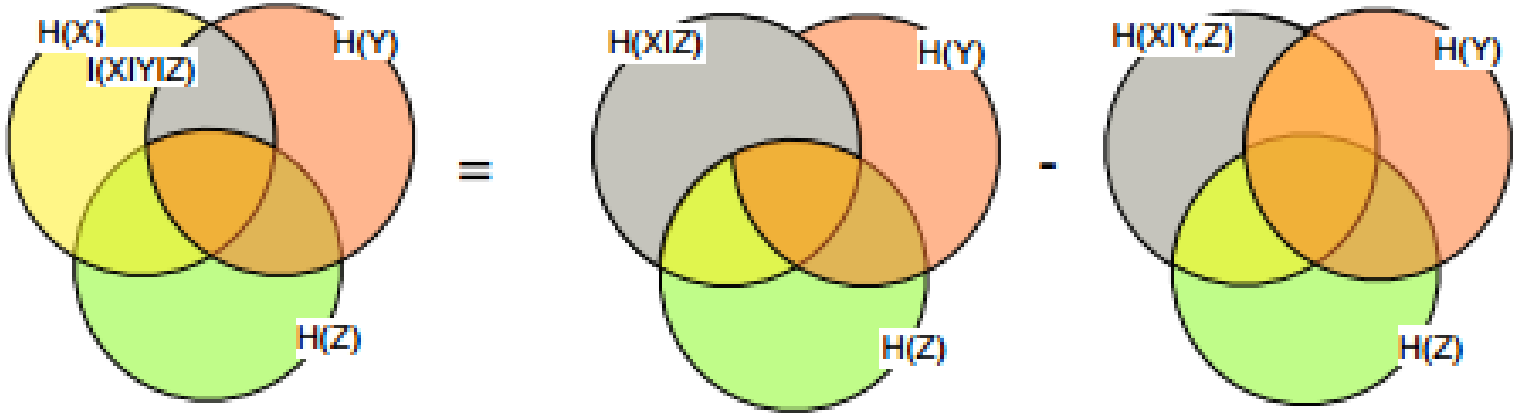
Conditional Mutual Information

Definition: The conditional mutual information between X and Y given Z is



$$I(X;Y|Z) := H(X|Z) - H(X|Y,Z)$$
$$= E_{p(x,y,z)} \log \frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)}$$

$$I(X;Y) = H(X) - H(X|Y)$$
$$= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$$
$$= E_{p(x,y)} \left[\log_2 \frac{p(X,Y)}{p(X)p(Y)} \right]$$



Mutual Information and Correlation

Correlation:

1. Correlation measures the linear relationship or monotonic relationship (e.g. Pearson's correlation or Spearman's correlation) between two variables, X and Y.

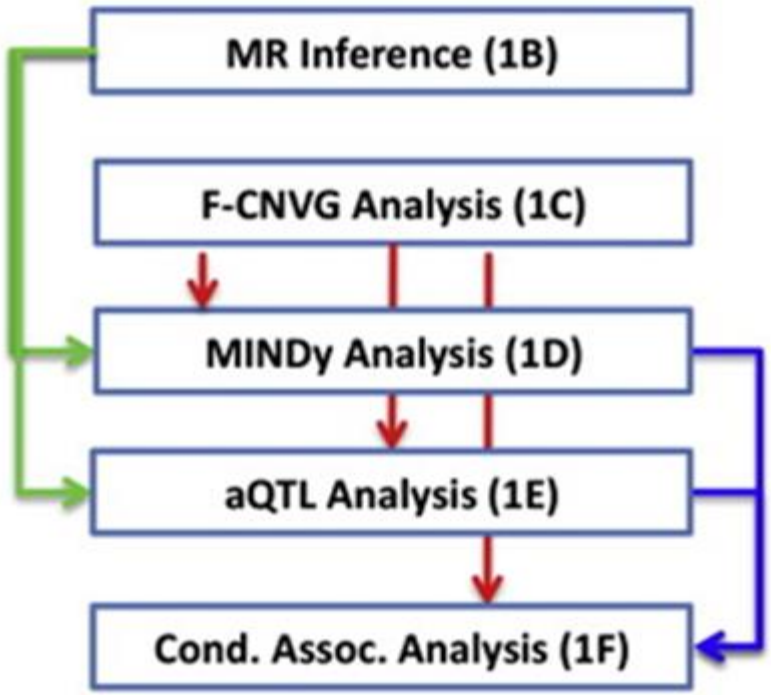
Mutual Information:

1. Mutual information is more general and measures the reduction of uncertainty in Y after observing X.
2. It is the KL distance between the joint density and the product of the individual densities.
3. So MI can measure non-monotonic relationships and other more complicated relationships.

DIGGIT

Methods / Process

DIGGIT: Overall Process



Overall flowchart of the DIGGIT pipeline.

Green: Use of MR Inference results

Red arrows: Use of F-CNVGs results

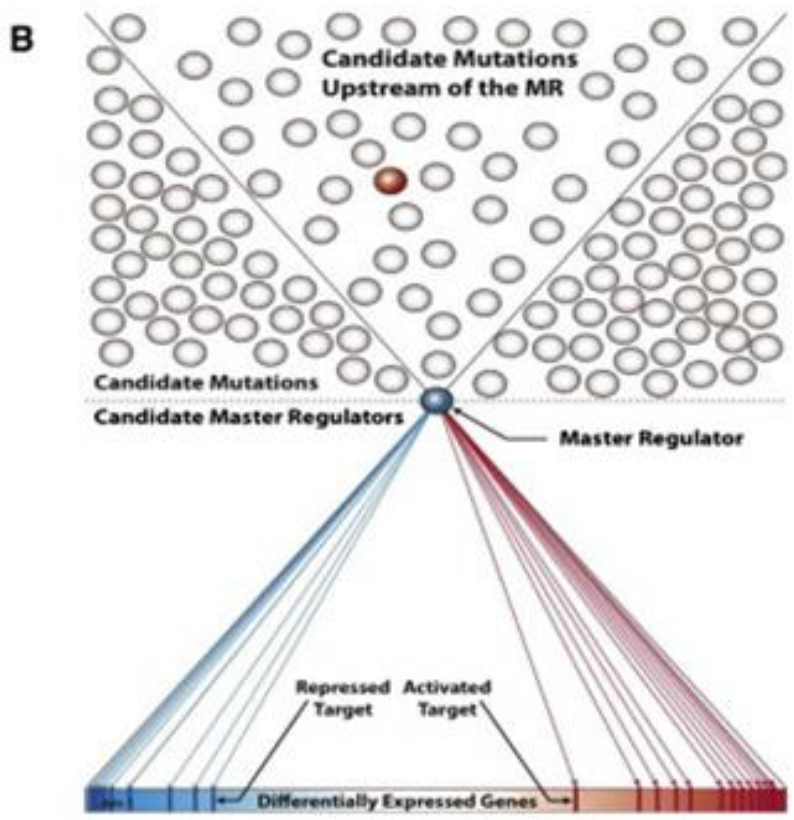
Blue arrows: MINDy/aQTL analysis results

1. 5-step pipeline process
2. Inputs:
   - ❑ Large set of Gene Expression Profiles (GEPD)
   - ❑ Sample matched Genetic Variant Profiles (GVPD)
   - ❑ Accurate and comprehensive repertoire of cell-context-specific molecular interactions (Interactome)
3. Output:
   - ❑ A p-value ranked list of candidate driver F-CNVGs.

Step-0: ARACNE

1. **ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks),** a novel algorithm, uses microarray expression profiles to reverse engineer human regulatory network.
2. Specifically designed to scale up to the **complexity of regulatory networks in mammalian cells**, yet general enough to address a wider range of network deconvolution problems.
3. This method uses an information theoretic approach (Mutual Information) to eliminate the vast majority of indirect interactions typically inferred by pairwise analysis.
4. On synthetic datasets, ARACNE achieves extremely low error rates and significantly outperforms established methods, such as Relevance Networks and Bayesian Networks.
5. DIGGIT uses ARACNE to reverse engineer the cellular network (Interactome) from GEPD

Step-1: MR Analysis



Objective:

Identify candidate MRs as TFs that activate over-expressed and repress under-expr genes.

Inputs:

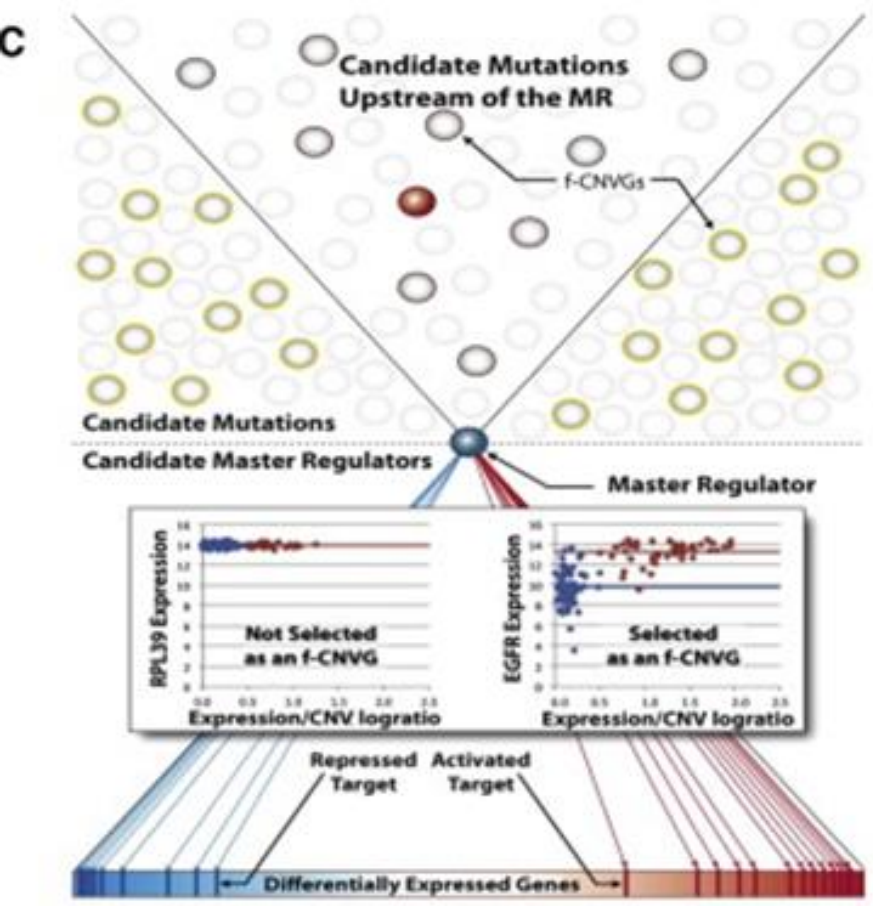
1. Context specific regulatory network (Interactome) rev-engineered from GEPD set
2. Gene expression signature of interest

Results:

Identified 6 MR genes - C/EBPβ, C/EBPδ, STAT3, BHLHB2, RUNX1, and FOSL2

1. Inferred using the MARINa algorithm.
2. One MR (blue circle) is represented in the panel.
3. Grey circles represent the repertoire of genetic alterations that may be associated with the phenotype
4. Those within the two diagonal lines (funnel) represent alterations in pathways upstream of the MR.
5. The red circle represents a bona fide causal driver alteration.

Step-2: F-CNVG Analysis



Objective:

Identify candidate functional CNVs (F-CNVGs).

Inputs:

1. GEPD & sample matched GVPD.

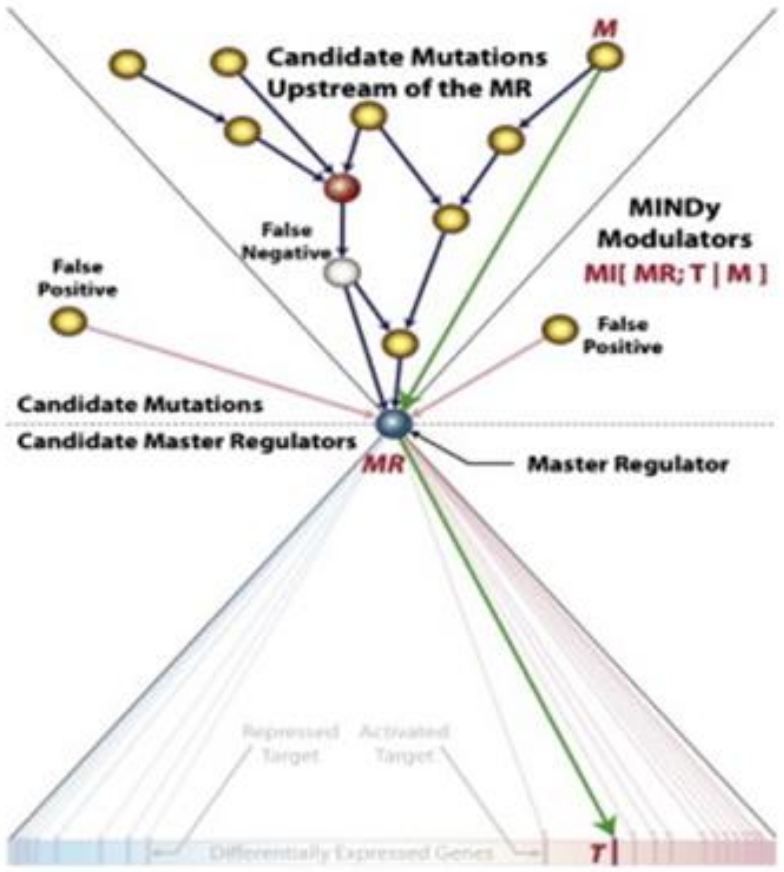
Results:

Identified **1,486** candidate F-CNVGs.

Inferred F-CNVGs included most genes previously reported as GBM drivers (14/18 > 88%).

1. F-CNVGs are determined by association analysis of copy number and gene expression.
2. Select copy-number alterations (CNVGs) whose ploidy is informative of gene expression as **candidate functional CNVs (F-CNVGs).**
3. Assessed based on **(1) mutual information (MI) between copy number and expression or (2) differential expression in wild-type (WT) versus amplified/deleted samples.**
4. Removes a large number of genes whose expression is not affected by ploidy.
5. The insert shows two examples: (a) an example of **no dependency between copy number and expression** and not selected as a candidate F-CNVG and (b) an example with **highly significant dependency** and thus selected as a candidate F-CNVG.

Step-3: MINDy Analysis



Objective:

Identify F-CNVGs that are candidate post-translational modulators of MR activity.

Inputs:

MR list(step 1) & F-CNVG list (step 2).

Output:

Generates a p value-ranked list of candidate F-CNVGs in pathways upstream of MR genes.

Results:

Identified **92** statistically significant candidate MES-MR modulators.

1. Use Conditional Mutual Information: Compute the cMI I[MR;T|M], where M is a candidate modulator gene and T is an ARACNe-inferred MR-target gene.
2. Blue arrows represent physical signal-transduction interactions upstream of the MR.
3. Green arrows represent one specific M → MR → T triplet tested by MINDy, as an example.
4. MINDy does not infer the blue arrows but only the fact that a protein is an upstream modulator of MR activity.

CMI in MINDy Analysis

cMI between an MR and its target (T), given a candidate modulator protein (M) is defined as:

$$cMI[MR;T|M] = \sum_{MR}\sum_{T}\sum_{M} p_{MR,T,M}(mr,t,m)\log\frac{p_M(m)p_{MR,T,M}(mr,t,m)}{p_{MR,M}(mr,m)p_{T,M}(t,m)}.$$

Under certain assumption, see (Wang et al., 2009), we can infer a modulatory effect of M on TF if the cMI is different from the MI, i.e.:

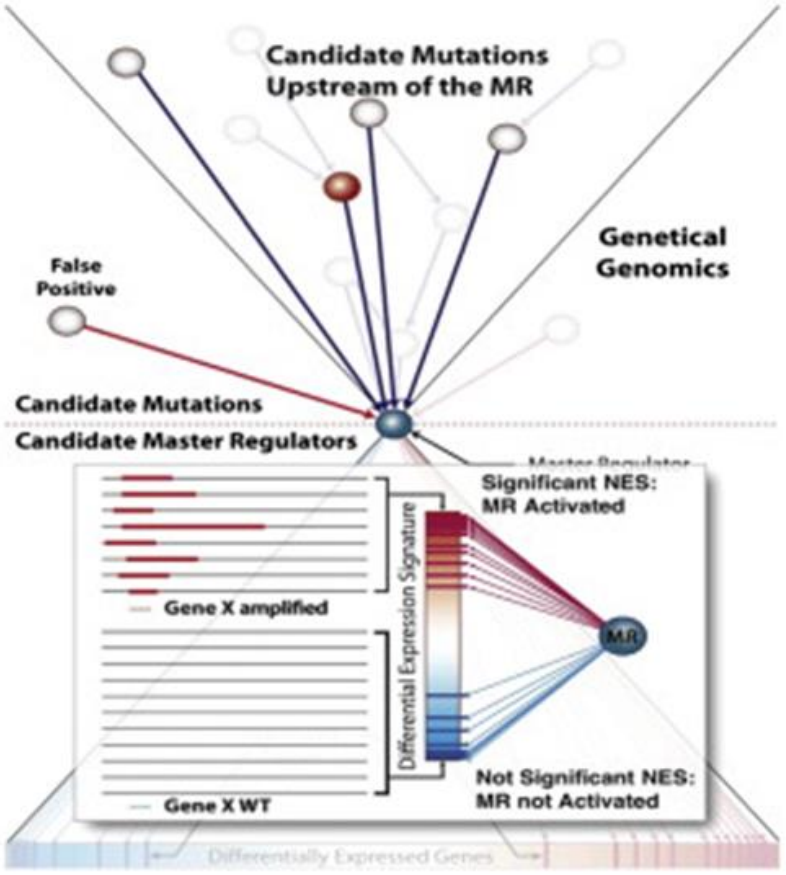
$$cMI[MR;T|M] - MI[MR;T] \neq 0.$$

This can be estimated more efficiently by measuring the change in mutual information after binning for high and low expression of the modulator *M*:

$$\Delta MI = MI\left[MR;T\middle|M_{high}\right] - MI[MR;T|M_{low}] \neq 0,$$

where M_{low} and M_{high} represent the $1/3^{rd}$ samples with the lowest and highest expression of the modulator M, respectively. This can be easily measured using the MI function defined in methods section. If the MI between a MR → T pair changes significantly as a function of a candidate modulator gene, then M is a candidate modulator of the MR activity.

Step-4: aQTL Analysis



Objective:

Identify F-CNVGs whose alterations cosegregate with aberrant MR activity.

Inputs:

MR list (step 1), F-CNVG list (step 2), GEPD data set, and the Interactome.

Output:

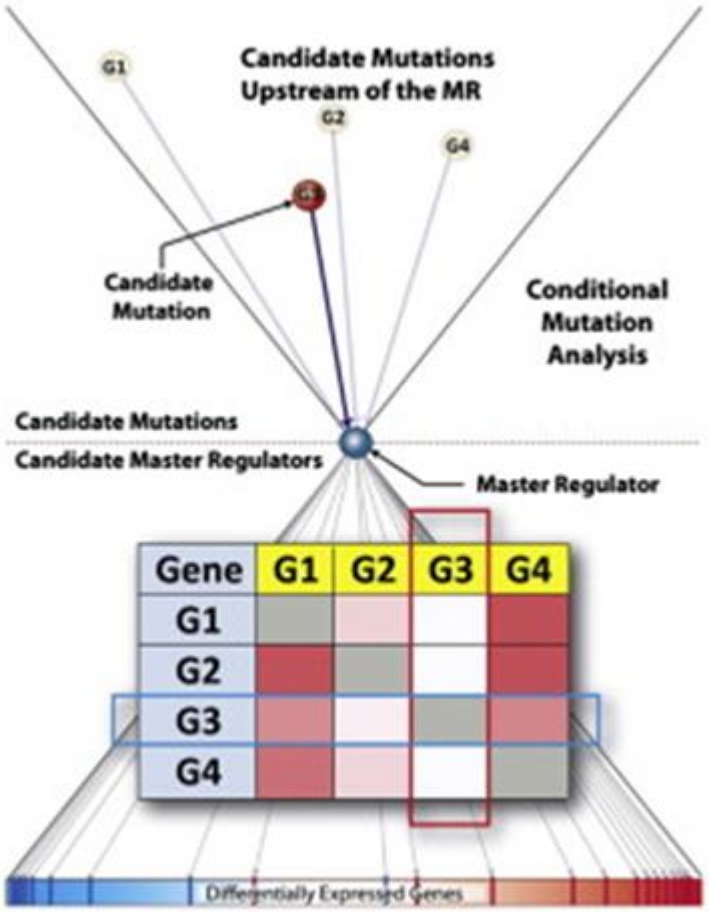
Generates a p value-ranked list of candidate F-CNVG-aQTL.

Results:

125 out of 1,486 F-CNVGs from step 2 were inferred as aQTLs.

1. Activity quantitative trait loci (aQTL) are inferred based on the statistical significance of the MI between copy number and MR activity.
2. Differential MR activity is inferred from their differential target expression, using a single-sample version of MARINa. Cosegregation computed (shown by the blue arrows).
3. The vertical gradient rectangle shows all genes sorted from the most over-expressed (red) to the most under-expressed (blue), when comparing samples with copy-number alterations in a gene (Gene X) (thick red lines) to WT samples (thin black lines).
4. If MR targets significantly cosegregate with the differential expression signature (i.e., if positively regulated and repressed MR targets, shown as red and blue bars, are over- and under-expressed, respectively, as shown), then Gene X alterations are likely to affect MR-activity.

Step-5: Conditional Association Analysis



Objective:

Identify F-CNVGs that abrogate all other associations with the phenotype (e.g., the MES-GBM subtype) when samples harboring their alterations are removed from the analysis.

Inputs:

MINDy/aQTL-prioritized F-CNVGs (steps 3/4), a phenotypic classifier, and GEPD data set

Output:

Generates a final p value-ranked list of candidate driver F-CNVGs

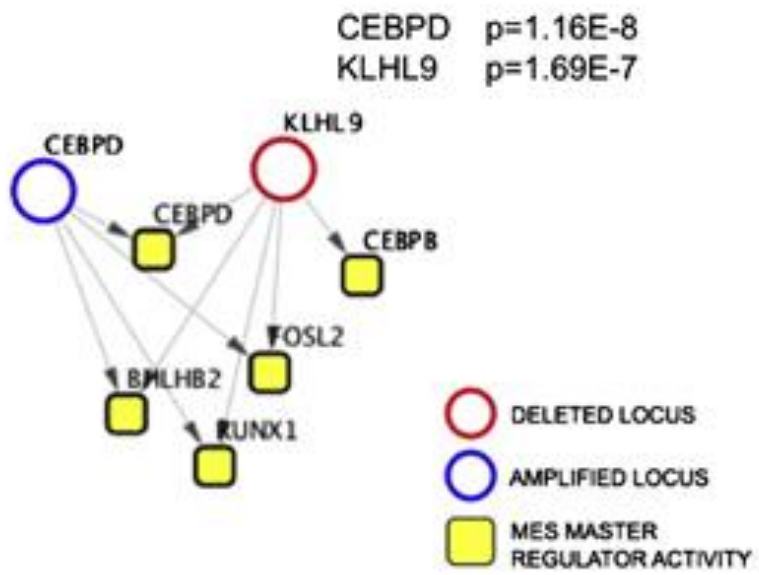
Results:

C/EBPδ and *KLHL9* abrogated asscіation of all other F-CNVGs, while remaining significant

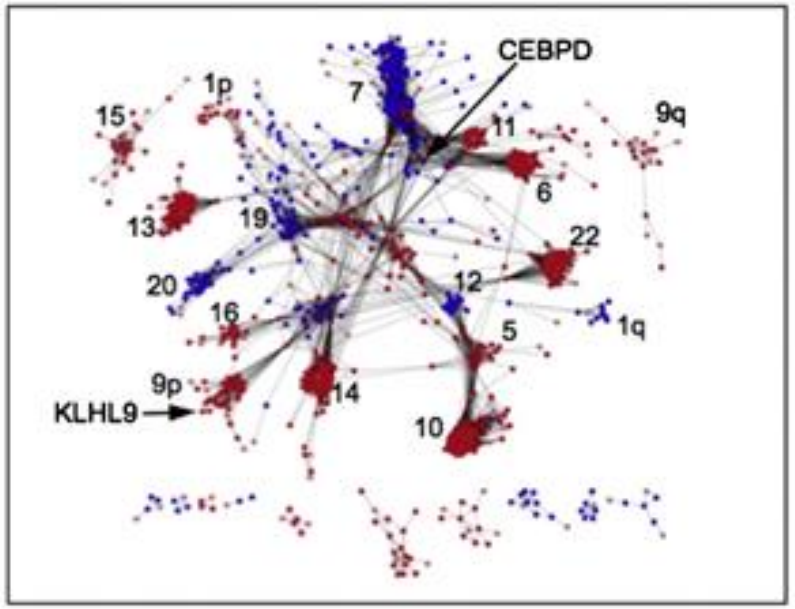
Conditional analysis discarded *CDKN2A*, a well-established tumor suppressor

1. Use conditional association analysis
2. Each cell shows the statistical significance of the association between the i-th gene (rows) and the phenotype of interest (as a heatmap), when considering only samples that have no alterations in the j-th gene (columns).
3. For instance, when conditioning on G3, no other gene is significantly associated with the subtype, whereas G3 is still significantly associated with the subtype when conditioning on G1, G2, or G4.
4. This suggests that G3 is a bona fide driver gene.

DIGGIT Integrative Analysis Infers Candidate MES-GBM Driver Mutations

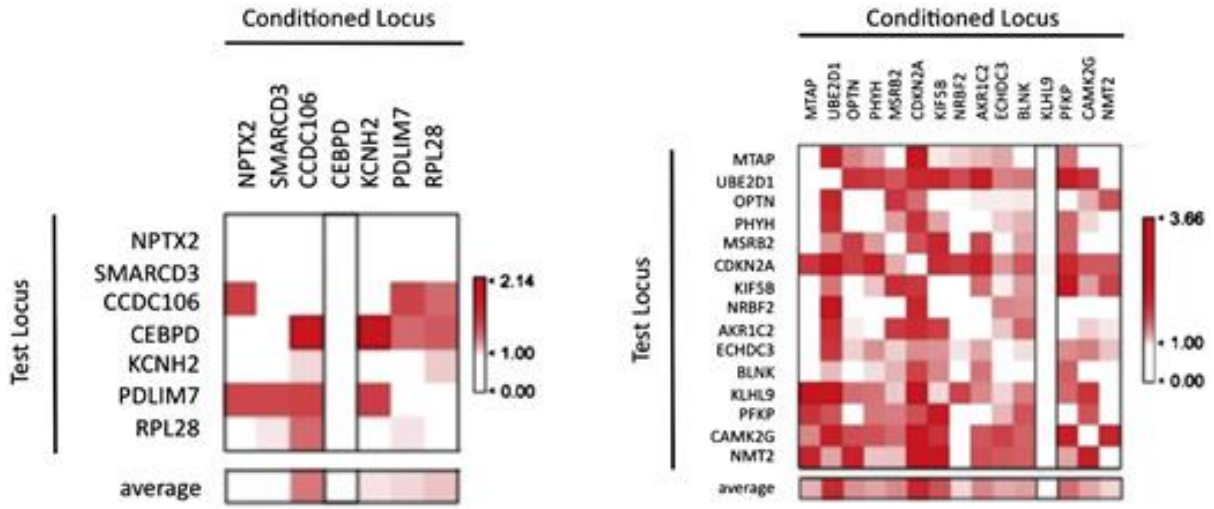


DIGGIT analysis of pathways upstream of MES-GBM MRs identifies **CEBPδ amplification and KLHL9 deletions** as candidate genetic determinants of the GBM-MES subtype. p values shown represent the integrated p value of the aQTL and MINDy steps.



Co-mutated F-CNVGs are shown as a network, with distance between connected nodes **inversely proportional to the statistical significance of their cosegregation**, as assessed by Fisher's exact test (FET). Only statistically significant pairs are shown (p = 0.05, corrected), with amplifications and deletions represented as blue and red nodes, respectively.

DIGGIT Integrative Analysis Infers Candidate MES-GBM Driver Mutations



Functional Mutation	Odds Ratio	pFET	# Poor Prognosis
$CEBPD^{amp}$; $KLHL9^{-/-}$	**>20.6***	**1.94E-3**	**17**
$CEBPD^{amp}$; $KLHL9^{WT}$	**9.50**	**9.90E-3**	**14**
$KLHL9^{-/-}$; $CEBPD^{WT}$	**5.47**	**2.12E-5**	**38**
$EGFR^{amp}$; $KLHL9^{WT}$,$CEBPD^{WT}$	1.35	0.227	**Total: 69/144, 48%**
$CDKN2A^{-}$; $KLHL9^{WT}$,$CEBPD^{WT}$	1.03	0.524	
$CDK4^{amp}$; $KLHL9^{WT}$,$CEBPD^{WT}$	0.631	0.910	
$PDGFRA^{amp}$; $KLHL9^{WT}$,$CEBPD^{WT}$	0.282	0.998	

*no good prognosis samples bear this genotype; OR reported is the maximum value detectable in the cohort

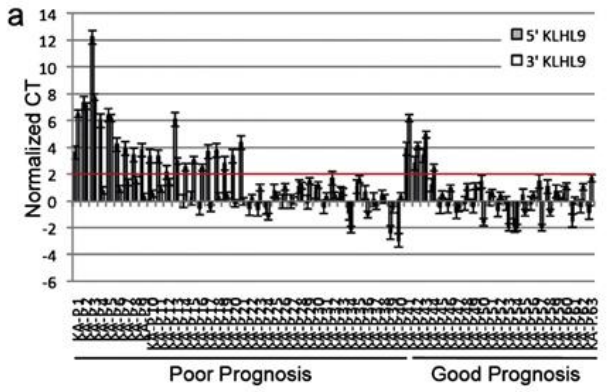
Conditional association analysis for the **two main co-segregating mutation clusters** identified by DIGGIT. Color scale in the matrix cell (i,j) represents the strength of association (−log10(p)) between the i-th F-CNVG (row) and the MES subtype, conditional to removing samples with alterations in the j-th F-CNVG (column).

Effect size of DIGGIT-inferred genetic determinants of the MES-GBM subtype. “Classical” GBM oncogenes are shown only as a reference, for comparison purposes.

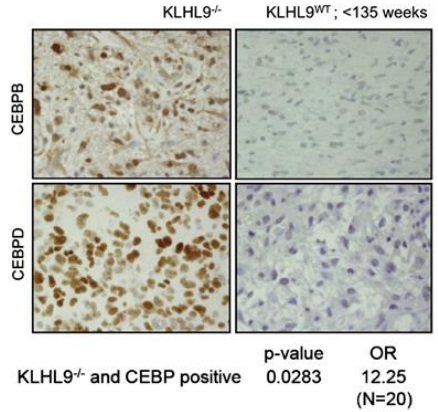
Key Takeaways

1. Only ***C/EBPδ*** and ***KLHL9*** abrogated association of all other F-CNVGs, while remaining significant when conditioning on other F-CNVGs
2. Conditional analysis discarded ***CDKN2A*** *(*a well-established tumor suppressor located proximally to *KLHL9)* as a candidate causal driver of MES-GBM.
3. **C/EBPδ amp** and **KLHL9 –/–** events account for 48% of TCGA MES-GBM samples
4. Along with independent deletions/mutations of NF1 covering an additional 8%, these may constitute the most common subtype drivers.

Association of KLHL9 Deletions Is Confirmed in an Independent Cohort

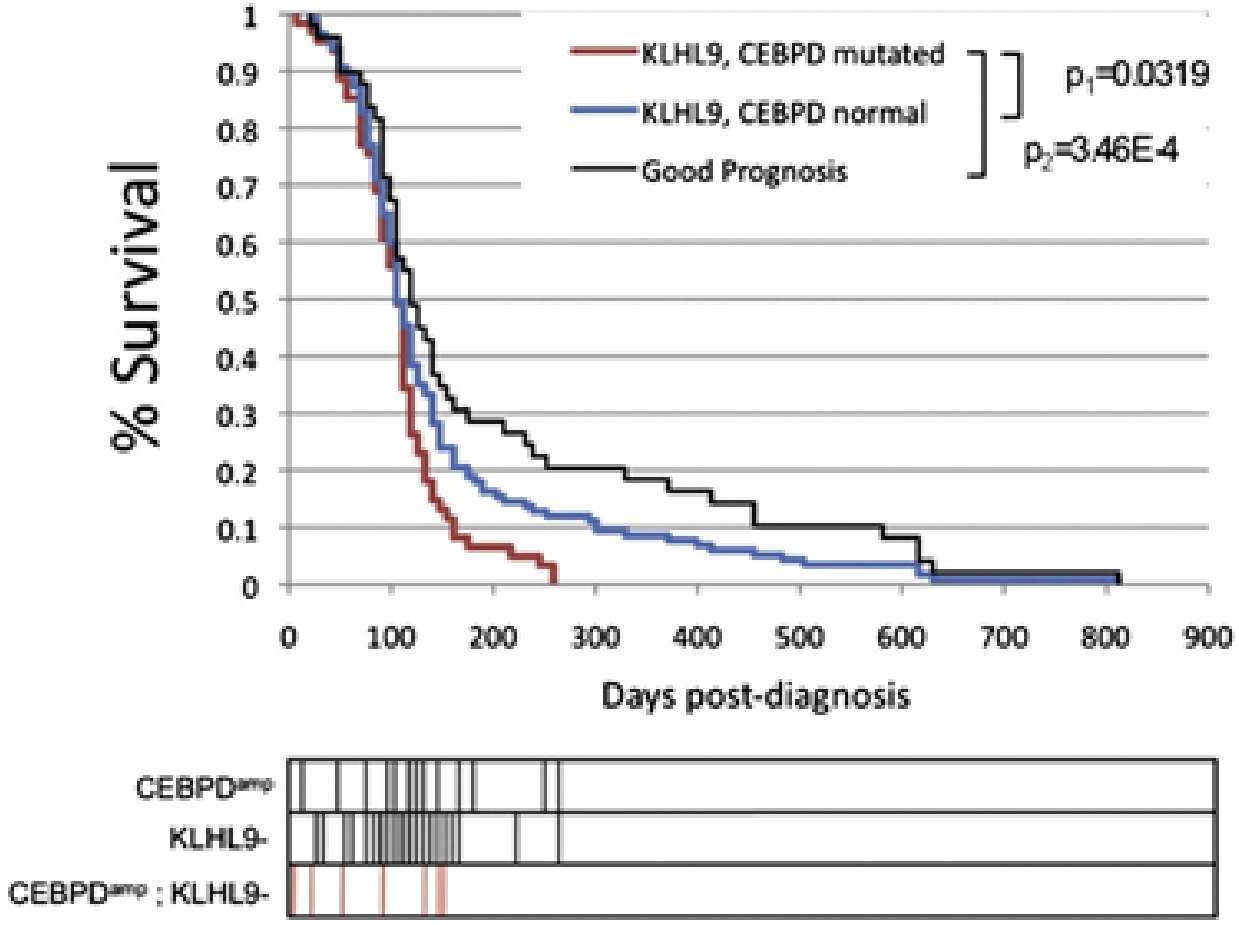


	$KLHL9^{-/-}$	Total Samples
Poor Prognosis	21	40
Good Prognosis	4	23
p-value	5.68×10^{-3}	



1. Tested whether association of *KLHL9* deletions with poor prognosis could be validated in an independent cohort.
2. Analyzed 63 FFPEs, representing 40 poor-prognosis (survival < 35 weeks) and 23 good-prognosis (survival > 130 weeks) GBM samples.
3. Quantitative genomic PCR revealed higher frequency of homozygous KLHL9 deletions in poor-prognosis (21/40) versus good-prognosis samples (4/23) ($p = 0.006$ by FET). Even higher frequency (>50%) than in TCGA samples (38%).
4. IHC staining of 10 $KLHL9^{-/-}$ and 10 $KLHL9^{WT}$ confirmed association with aberrant C/EBPβ and C/EBPδ protein expression in vivo (odds ratio 12.25, $p = 0.028$).
5. **Confirms $KLHL9^{-/-}$ events as poor-prognosis biomarkers** and their association with aberrant MES-MR activity in vivo. No *KLHL9* missense or nonsense mutations were detected.

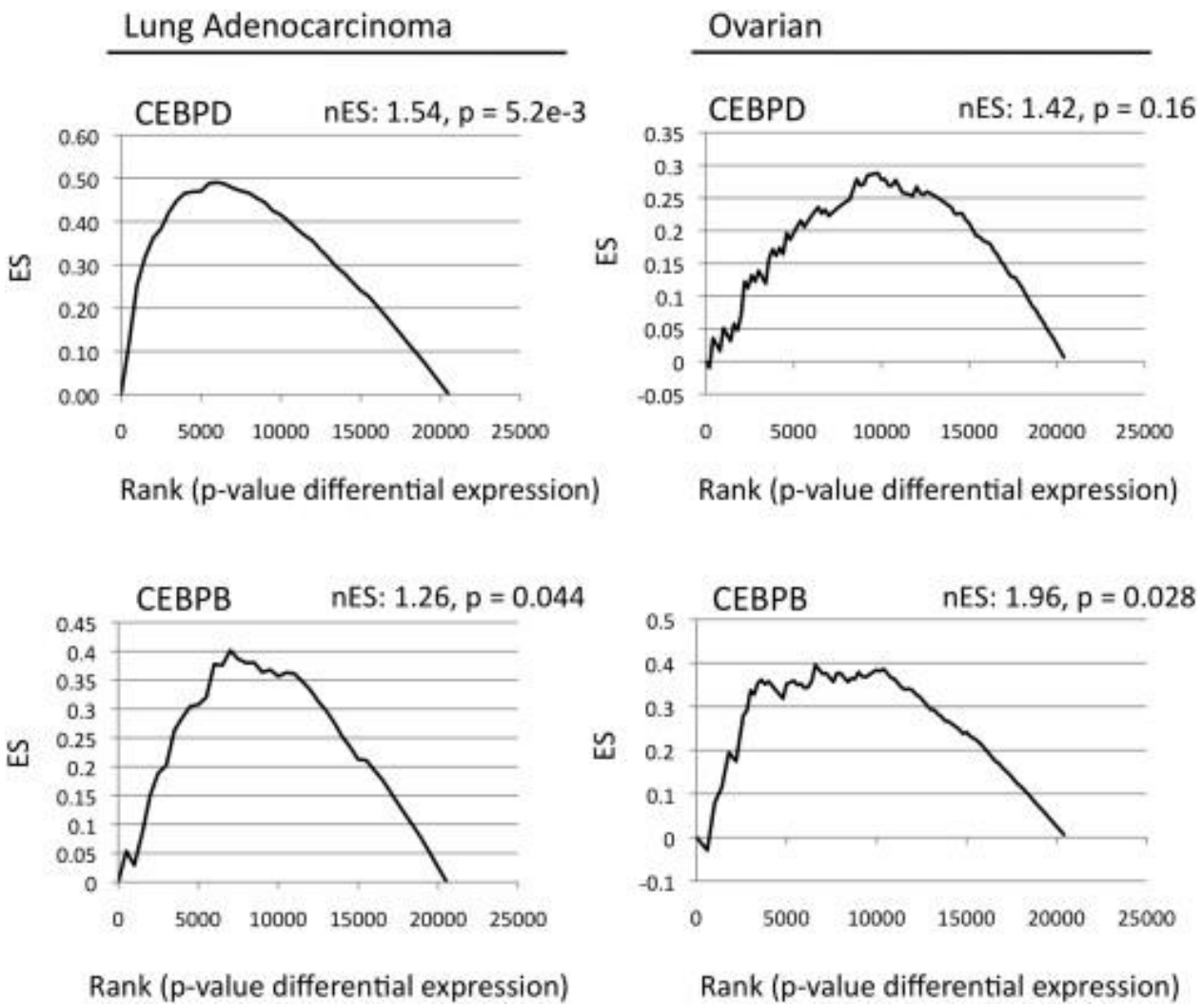
Association of KLHL9 Deletions Is Confirmed in an Independent Cohort



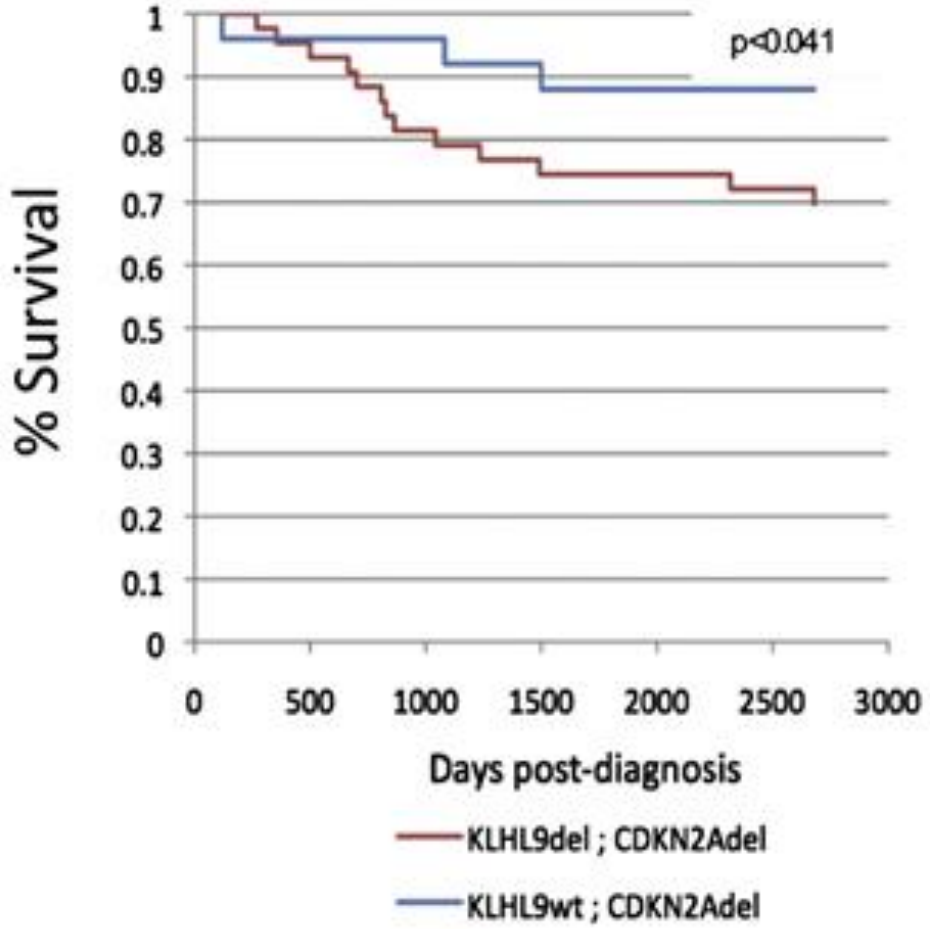
1. Kaplan-Meier analysis of GBM samples in TCGA.
2. Patients with KLHL9$^{-/-}$ and C/EBPδ^{Amp} events are shown as a red curve
3. Proneural subtype patients are shown as a black curve
4. KLHL9^{WT}/CEBPδ^{WT} samples are shown as a blue curve
5. Kaplan-Meier p values are shown, including p1 (red versus blue) and p2 (red versus black).
6. Survival for patients with each specific genotype is shown as vertical bars below the plot.

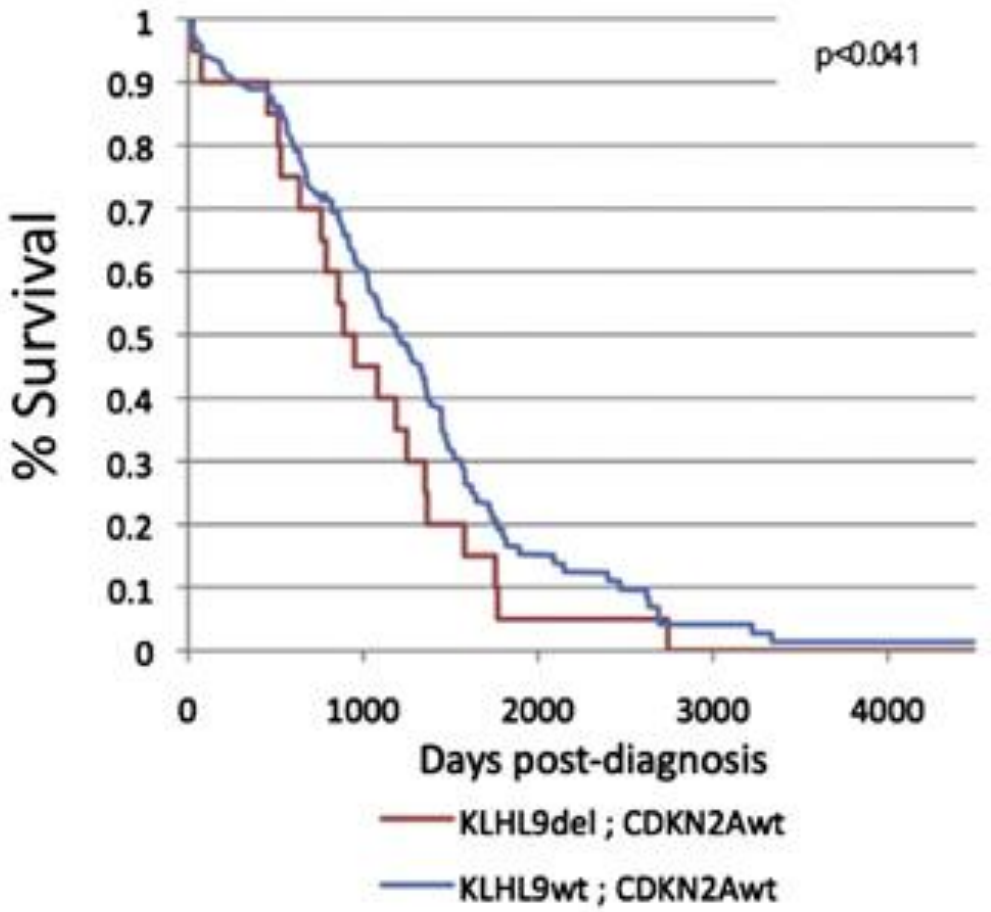
C/EBPδ and KLHL9 Alterations Are Predictive of Poor Prognosis in Multiple Tumors

1. Assessed whether C/EBPδ^{Amp} and KLHL9$^{-/-}$ events may be **predictive of poor prognosis in GBM and other tumors**.
2. In GBM, Kaplan-Meier analysis revealed significantly worse prognosis for patients **harboring C/EBPδ^{Amp} and KLHL9$^{-/-}$** alterations, compared to either good-prognosis or C/EBPδ^{WT}/KLHL9^{WT} patients.
3. None of the patients with these alterations survived longer than 36 weeks post-diagnosis, and patients harboring both events had the worst overall prognosis, suggesting a **cooperative effect**.
4. Thus, C/EBPδ^{Amp} and KLHL9$^{-/-}$ **represent genetic biomarkers of poor prognosis**, independent of subtype classification.
5. Kaplan-Meier analysis revealed that KLHL9 homozygous deletions and missense/nonsense mutations are associated with the **worst prognosis also in Lung (LuAd) and Ovarian (OvCa) adenocarcinomas** independent of CDKN2A status.
6. Gene set enrichment analysis (GSEA) confirmed aberrant C/EBPβ and/or C/EBPδ activity in KLHL9$^{-/-}$ samples, suggesting a **possible pan-cancer role of KLHL9 deletions via aberrant C/EBP activity**.



1. Enrichment analysis of CEBPB and CEBPD ARACNe-inferred targets in genes differentially expressed in $KLHL9^{-/-}$ versus $KLHL9^{WT}$ samples.
2. Results for both lung adenocarcinoma (LuAd) and ovarian cancer (OV) are shown.
3. This analysis confirms that C/EBP protein activity is aberrantly increased by loss of KLHL9 function in multiple tumor types.





Kaplan-Meier analysis of the association between KLHL9$^{-/-}$ alterations and poor prognosis in lung and serous ovarian adenocarcinoma, respectively. Analysis of inferred differential activity of C/EBPβ and C/EBPδ in KLHL9$^{-/-}$ samples.

Ectopic (Unusual) *KLHL9* Expression in GBM Cells Abrogates (stops) C/EBPβ and C/EBPδ Abundance

1. To **mechanistically elucidate** KLHL9-mediated regulation of established MES-MRs (C/EBPβ, C/EBPδ, and STAT3), **rescued KLHL9 expression** in homozygously deleted cells.
2. Used cell lines SF210 and SF763 cells labeled as $KLHL9^{-/-}$;$CDKN2A^{-/-}$;C/EBP^{WT}.
3. **RNA-seq profiling revealed significant differential expression** of ARACNe-inferred C/EBPβ and C/EBPδ targets by GSEA compared to controls.
4. Involved significant **down-regulation of established MES markers**: CHI3L1/YKL40, LIF, FOSL2, ACTA2, and FN1.
5. Observed significant reduction in C/EBPδ and more modest decrease in C/EBPβ protein levels.
6. Levels of phospho-STAT3, representing the transcriptionally active isoform, were also reduced.
7. **Exogenous expression of P16/INK4A (CDKN2A) had no effect** on either C/EBPβ or C/EBPδ protein expression or on MES signature genes.
8. Results show that rescue of KLHL9 expression **collapses the MES-GBM signature** by downregulating C/EBPβ and C/EBPδ at the protein level.

Proteasomal Degradation of C/EBPb and C/EBPd Depends on KLHL9-Mediated Polyubiquitylation

1. KLHL9's has a putative function as an adaptor of Cul3-based E3 ubiquitin ligase.
2. Tested KLHL9's role in mediating polyubiquitylation-dependent proteasomal degradation of C/EBPb and C/EBPd.
3. Measured degradation and relative half-life of C/EBPb and C/EBPd following rescue of KLHL9 expression in SF210.
4. MG-132-mediated proteasome inhibition abrogated C/EBPb and C/EBPd degradation, confirming that KLHL9 is required for their proteasomal processing.

More Mechanistic Insights

1. KLHL9 Mediates Polyubiquitylation of C/EBPb and C/EBPd Isoforms
   - Confirmed that proteasomal degradation of C/EBPs depends on KLHL9-mediated interaction with the CUL3 E3 ligase complex
   - Confirmed that KLHL9-mediated C/EBP regulation depends on a functional KLHL9-CUL3 E3 ligase complex
2. KLHL9 Expression Delays Exit from S Phase in Glioma Cells
   - Confirmed that rescue of KLHL9 expression delays the cell cycle by imposing a late S/G2 checkpoint.
3. KLHL9 Expression in $KLHL9^{-/-}$ Patient-Derived GBM Tumors Reduces Growth in Orthotopic Xenografts
   - Experiments show that in vitro cell-cycle-dependent reduction in proliferative potential, induced by ectopic KLHL9 expression in human cell cultures, is recapitulated in vivo and induces retardation in tumor growth.

Unbiased Inference of Driver Alterations in BRCA and AD

1. **Analysis of sample-matched CNV/expression data from the TCGA breast cancer (BRCA) cohort.**
   - Compiled a list of 25 alterations from a literature search of validated CNV alterations linked to BRCA tumorigenesis
   - Final step (Conditional Association Analysis) yielded 35 F-CNVGs
   - Of these, 19 (76%) could be matched in the 25-gene literature compiled list
   - Only 5 of these were statistically significant by Genome Wide Association Studies (GWAS)
2. **Analysis of sample-matched SNP/expression data from a recent integrative study of Alzheimer's disease.**
   - DIGGIT identified 13 F-SNPs significant by conditional association analysis.
   - Among these, TYROBP was ranked 1st ($p = 4.2 \times 10^{-47}$), achieving higher significance than even APOE, ranked 9th ($p = 2.0 \times 10^{-21}$)