QIAGEN

Application note

Understanding Breast Cancer Subtypes by Jointly Interpreting Tumor Genomes and Transcriptomes

Jean-Noël Billaud, Ph.D. Principal Scientist, External Scientific Relations, RNASeq Solution, QIAGEN Bioinformatics

Identifying differences among tumor subtypes requires a multifaceted approach that integrates information about genotypes and gene expression patterns and highlights the impact of these differences within the context of healthy and disease physiology. Ingenuity® Variant Analysis™ and Ingenuity Pathway Analysis (IPA®) are web-based applications that enable the rapid interpretation and integration of human genome and transcriptome data to generate novel hypotheses about the regulation and effects of change patterns observed in analyzed data. Powerful, interactive tools and statistically robust methods for studying tumors, probands, kindreds, and large case-control cohorts leverage the Ingenuity Knowledge Base to identify experimentally relevant candidate variants and interrogate canonical pathways and subnetworks, thereby inferring potential upstream activity and simulating downstream effects.

Introduction

Deep nucleic acid sequencing can help personalize healthcare by tracing observed variation in disease incidence and outcome to molecular differences among people and tissues. Such efforts can only reliably inform treatment via integrative analysis of complex patient-specific data, leveraging functional knowledge of proteins, drugs, metabolites, and other associated molecules. Tumors — being genetically noisy, transcriptionally distinctive, and medically urgent — are a key proving ground for relevant methods to jointly interpret genomes and transcriptomes, in order to trace tissuespecific physiology to specific variants that can be effectively screened or targeted in treatment.

This Application Note shows how jointly analyzing tumor-specific genotypes and gene expression can shed light on key differences among tumor subtypes. Pairing 2 tools from QIAGEN's Ingenuity Systems — Variant Analysis for interpreting human genome data and IPA for transcriptome data — differences were traced between breast tumors that spread quickly (Claudin-low) versus slowly (luminal) to sequence variation that likely governs epithelial-to-mesenchymal transition (EMT). EMT is a complex de-differentiation program, key to early embryonic development, in which epithelial cells lose polarity and epithelial surface markers, and start expressing mesenchymal markers typical of fibroblasts and motile cells (1). Tumor lineages that undergo EMT can more readily migrate and invade distal organs.

Most Claudin-low breast tumors are metaplastic/medullary-differentiated triple-negative (ESR1, PGR, ERBB2) invasive ductal carcinomas with poor prognosis that express few or no markers for luminal differentiation (CD24, CD326/EpCam, MUC1, GATA3, etc.) or tight junctions (CLDN3/4/7 (claudin), CDH1 (E-cadherin)), but strongly express markers of immune response and, most strikingly, EMT (2).

By contrast, luminal breast tumors (subtypes A and B) are less stem cell-like, and strongly express ESR1, KRT8, and KRT18(2). Among intrinsic subtypes, luminal tumors tend to have

the best prognoses and respond well to endocrine therapy.

To identify candidate drivers of EMT, deep RNAseq data from 5 representative cell lines of each subtype were interpreted. Using Variant Analysis and IPA, cell line-specific data were systematically annotated and compared, leveraging the Ingenuity Knowledge Base of millions of curated and predicted relationships among variants, genes, proteins, complexes, cells, tissues, metabolites, drugs, and diseases. With Variant Analysis, putative genomic variants were filtered by subtype, distribution, and likely functional relevance to quickly and confidently shortlist biologically plausible genomic candidates. Chief among the resulting variant candidates was a novel missense in HRAS, recurrently called in Claudin-low subtypes, that likely confers EMT-activating gain-of-function. The functional and predictive algorithms of IPA suggest that this variant may drive strong enrichment of pro-metastatic transcripts in Claudin-low samples, consistent with HRAS activation of causal pathways driven by transcription factors SNAI1 and TWIST, and molecular markers of EMT, such as MMP2 (matrix metalloproteinase-2).

Methods

RNAseq data (~75 million paired-end Illumina® HiSeq® 2000 reads, generated by Expression Analysis, Durham, NC) from each of 5 luminal (BT-549, HS578T, MDA-MB-231, SUM1315-PT, SUM159PT) and 5 Claudin-low (HCC1428, MCF-7, MDA-MB-361, T47-D, ZR-751) cell lines were filtered, aligned by Bowtie to human reference transcriptome (UCSC KnownGene) or, for initially unmapped reads, aligned by BWA to human reference genome (Hg19), and called using GATK. Transcripts were assembled, and relative abundances estimated and compared, via Cufflinks. The intensity (FPKM) for each transcript across the cell lines in each group was averaged, and the maximum isoform-level change per gene (fold change cutoff 2, $p < 0.05$, FPKM cutoff 100) was set. All upstream tools were run via Golden Helix (Bozeman, MT), yielding VCF files that were uploaded into the Variant Analysis platform, in order to identify candidate subtype-distinctive genomic drivers by comparative filtering (see below). IPA was then used to assess relevance of differentially expressed genes (isoforms) and identify the putative causal variant in EMT.

Results

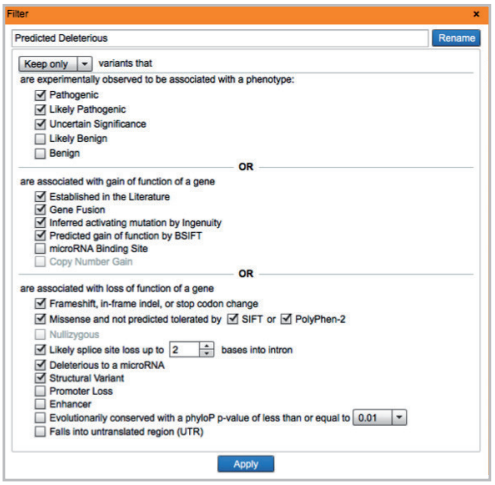
Variant Analysis

Variant Analysis is a secure (HIPAA- and Safe Harbor-compliant) web platform for annotating and comparing comprehensively sequenced human genomes to quickly shortlist candidate variants in studies of matched or unmatched (as here) tumors, disease kindreds, single- or multi-proband sets, or large case-control cohorts. A few simple questions are asked at the start of analysis, after which the platform sensibly parameterizes filters for finding credibly rare, appropriately functionally suspect variants, based upon study design, focus, and assumptions. After analysis, Findings can be reviewed and filters tweaked on-the-fly to dynamically update shortlists, in order to leverage new insights or assumptions. The data sharing function allows for easy and secure transfer

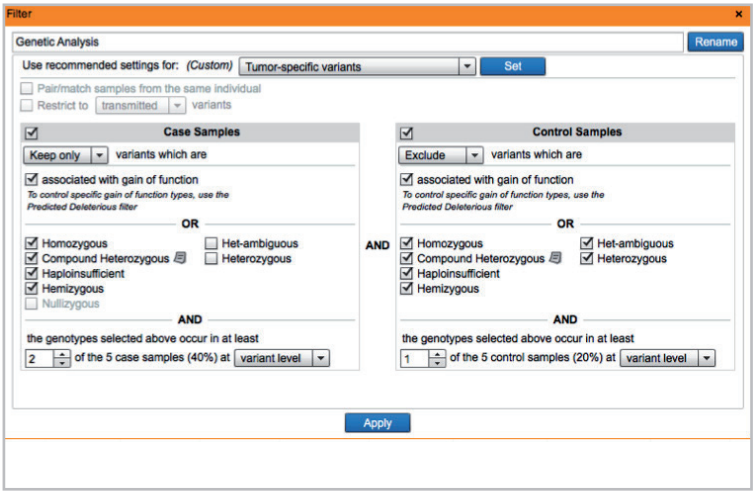
of results and associated analysis settings. In a given tumor cell line, few of the hundreds of thousands of distinctive genetic variants found are likely to drive its distinctive physiology. Spotting such likely drivers requires sensible filters to accurately answer 3 key questions about each putative variant: Is it real? How common is it among other tumors and in the world at large? And how might it affect physiology, through gene product sequence or/and expression? Variant Analysis uses a sensibly default-configured, yet customizable series of filters to answer these questions and shortlist candidate variants, genes, and pathways.

Figure 1. Screenshots from Variant Analysis depicting the range of settings available for **A** predicted deleterious and **B** genetic analysis filters. All filters are highly configurable to match study design, focus, and assumptions.

A



B



Quality, credibility, and rarity filters

Starting with 387,662 putative non-reference variants, each called in at least 1 studied cell line, a confidence filter (invoking site-specific variant call quality and read depth) was used to narrow the list down to 194,352 likely real variants. A credibility filter was then used to exclude variants in genome segments that appear hypervariable in short read-sequenced healthy genomes; this filter removes real but harmless variants that are densely or sparsely distributed in genes of average or extreme length, along with spurious variants that reflect ancient segmental duplications missing from the reference genome. Finally, a rarity filter was used to reduce the list to 96,341 variants with estimated allele frequencies below 0.1% among non-cancerous genomes published by the 1000 Genomes Project, Complete Genomics, and (in exome-targeted genome segments) the NHLBI Exome Variant Server.

Predicted deleterious, genetic analysis, and cancer driver variants filters

Having filtered variants by likely validity and general rarity, the Ingenuity Knowledge Base was then leveraged to focus on such variants plausibly relevant to tumor subtype physiology. First, typical functionally suspect variants were identified using the predicted deleterious filter (Figure 1A) to pinpoint variants that either (a) fit ACMG criteria for classification as pathogenic or of uncertain significance (including likely pathogenic), (b) grossly alter RefSeq-annotated protein sequence (nonsense, read-through, frameshift, whole-codon indel, or likely splice-altering up to 2b into intron), (c) otherwise alter protein sequence and likely harm protein function (missense predicted to be harmful by SIFT or PolyPhen2), (d) alter an annotated miRNA or miRNA binding site, (e) reportedly confer gain-of-function in QIAGEN-curated, peer-

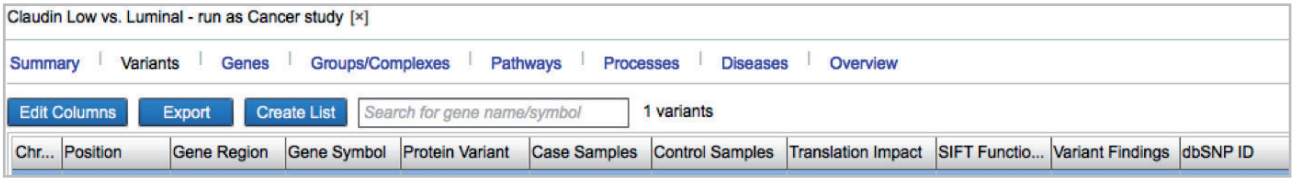


Figure 2. Screenshot of the results table view in Variant Analysis, showing the 1 variant resulting from the cancer driven variants filter. The table can be customized to display the information columns most relevant to the research question.

reviewed literature, (f) are predicted to confer gain-of-function by the BSIFT algorithm or the protein interaction evidence-based algorithm of Ingenuity, (g) are predicted to yield a gene fusion, or (h) alter a site strongly conserved (PhyloP ≤10-6). Next, a genetic analysis filter (Figure 1B) was applied to find such genes enriched for functionally suspect variants specifically found in Claudin-low cell lines. The filter was configured to keep only genes that harbored variants found in at least 2 Claudin-low cell lines, but in no luminal cell lines. This narrowed the list to 3,079 genes jointly called with 5,613 putatively real, generally functionally suspect, plausibly Claudin-low distinctive variants. Next, a cancer driver variants filter was applied to find variants among the foregoing likely to be generally relevant to tumor emergence or spread. These included variants in genes that (a) have cancer-related knockout phenotypes, (b) are appropriately implicated in cancer-relevant cellular processes, such as apoptosis, or in cancer-relevant pathways, (c) are targeted by cancer therapies, (d) are reportedly implicated in cancer(s) in QIAGEN-curated, peer-reviewed publications, (e) are reported or predicted to affect regulation of cancer-relevant protein interactions, (f) are reported to recur among at least 0.1% of tumor genomes published by COSMIC or TCGA, or (g) are reportedly implicated directly in breast cancer. Last, a biological context filter (named EMT-relevant) was used to find variants that fit all the foregoing criteria, and in genes specifically implicated in the key process — epithelial to mesenchymal transition (as searched in the Ingenuity Knowledge Base) — that distinguishes the 2 studied tumor subtypes. The search could have been expanded to include variants in genes directly implicated in other similar phenotypes, and/or reported as direct (1-hop, in Variant Analysis terminology) or indirect (2-hop) upstream regulators or downstream regulatory targets of such genes. The narrowly parameterized filter cascade zeroed in on 1 variant, a missense (p.G12D) in the EMT-implicated gene HRAS, that has been reported in many other aggressive tumors (3).

Assessing shortlisted candidates

Variant Analysis lists Findings in interactive tables and other simple graphical objects (such as gene-gene and gene-phenotype networks). The broadly useful variant table (Figure 2) summarizes Findings on shortlisted variants, including chromosomal position, gene(s) harboring the variant, sample-specific called genotypes (in a simple, easily scanned graphical summary that can be moused-over for more information), which samples hold the variant, the associated gene symbol(s), zygosity (including copy number information), predicted effects on gene product sequence and/or function, dbSNP ID (if known), and so forth. The table can be customized by adding, removing, or rearranging data fields, including ancillary custom annotation (uploaded via .bed files). Doubleclicking any variant shows further infor-

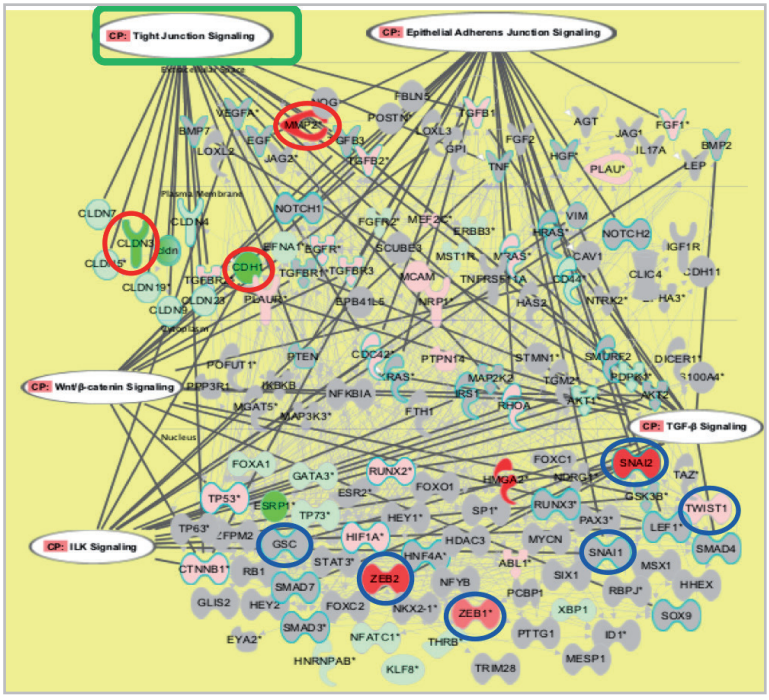


Figure 3. Network map generated by IPA depicting the genes involved in EMT, shaped by gene function and positioned by subcellular location. Genes are color-coded (red or green for up- and ownregulation,respectively) for the isoform expression of Claudin-low vs. luminalcontrol group. Red circles highlight known cancermarker MMP2 (upregulated) and CDH1, CLDN3 (down-regulated), while blue circles indicate key transcription factors of EMT (SNAI1/2, TWIST1, GSC, and ZEB1/2) many of which are up-regulated.Solid lines represent biological interactions of select genes within specific canonical pathways contributing to EMT.

mation, including detailed background information on gene(s) and a detailed, interactive graphical summary of Ingenuity Knowledge Base evidence linking the variant to any studied phenotype. In this study, HRAS p.G12D, which is predicted to activate protein function, was highlighted as a potential causal variant, as called in 2 of the 5 Claudin-low samples. HRAS protein, in turn, is implicated in EMT in the Ingenuity Knowledge Base (3).

IPA

IPA draws directly from the massive and high-quality literature-derived information within the Ingenuity Knowledge Base, offering interactive and visual tools to better understand the hidden biology in complex 'omics datasets at multiple levels. IPA provides insight into the molecular and chemical interactions, cellular phenotypes, and disease processes of a given system.

Visualizing biological networks and canonical pathways

Using the visual tools of IPA, the interactions within the molecular network of EMT are displayed. Figure 3 represents a network of the most important known molecular players of EMT, displaying their connections, their expression level, and their sub-cellular localizations. Specifically, MMP2, CDH1 (E-cadherin), and Claudins are highlighted (red circle) and key transcription factors in EMT (SNAI1, SNAI2, TWIST1, ZEB1, etc.) are indicated by blue circles. Genes are color-coded for up- or down-regulation (red and green, respectively), based upon the input cut-off criteria of the analysis. MMP2 and CDH1 show changes in regulation (up- and down-regulated, respectively) in the Claudin-low samples compared to the luminal control, as expected given their known association with cancer-related functions.

MMP2 is an important protease for extracellular matrix degradation, tumor invasion, and metastasis in malignant tumors. An up-regulated expression of MMP2 is associated with poor prognosis (e.g., prostate cancer (4)). A breakdown in the tight junction signaling pathway is a hallmark of EMT, with down-regulation of CDH1 expression, a key component of this system, associated with invasive and metastatic cancers (5). HRAS appears in gray (to the right of CDH1 in Figure 3) as it did not reach the cut-off applied for this original gene expression

analysis. Greater functional information can be drawn by linking up- or down-regulated genes to their related pathways and biological processes. Using the overlay feature, the most relevant canonical pathways are displayed over this EMT network. In particular, tight junction signaling appears as one of the most significant canonical pathways represented by the molecular network, as calculated by the Fisher's exact test, right-tailed. Tight junction strands are linear co-polymers of occludin, claudins, and JAMs, and are located at the uppermost portion of the lateral plasma membrane, where claudins appear to be involved in the homophilic and/or heterophilic interactions implicated in firm adhesions (6). Through a series of complex enzymatic cascades, tight junction assembly and apicobasal polarity are promoted and/or maintained. Under different stimuli (TNFA, TGFB, or others), the stability of the tight junction is perturbed (6). RhoA signaling (up-regulated in the Claudin-low subtype), regulates actin-based motility in high-density epithelial cells by inhibiting GEFH1, which in turn influences cell migration and cell cycle progression (7).

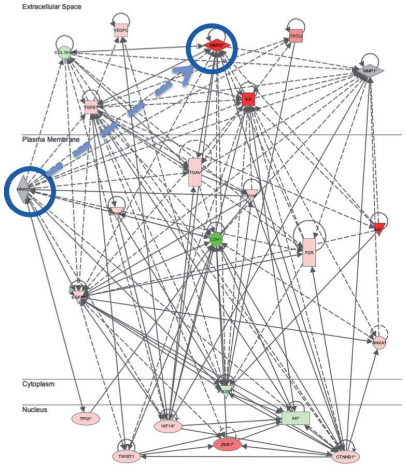


Figure 4. MAP depicting the network of relevant genes involved in metastasis. Shapes indicate gene function, placement is by sub-cellular location, and color indicates isoform expression (red: up-regulated, green: down-regulated) of Claudin-low versus luminal control group. Lines link genes with up- or downstream relationships as supported by data in the Ingenuity Knowledge Base. Solid or dotted edges indicate direct or indirect relationships, respectively. Direct interactions (chemical modifications) require that 2 molecules make direct physical contact with each other, as opposed to indirect interactions that do not have this requirement. The dotted line highlighted in blue indicates a relationship between HRAS and MMP2, supported by 4 references from the Ingenuity Knowledge Base.

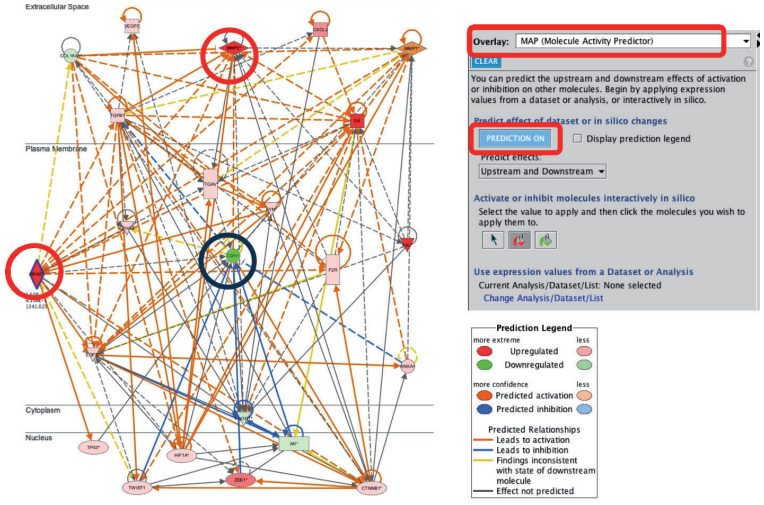


Figure 5. Network of genes involved in metastasis after applying the Molecular Activity Predictor function of IPA. Information relating to a gain in function of HRAS was used to simulate predicted directional consequences on the metastasis network. In silico interrogation of subnetworks and canonical pathways represented in the Ingenuity Knowledge Base support the upstream relationship of the HRAS causal variant to an increase in MMP2 activity and other expression pattern changes in the analyzed dataset.

Linking genetic variance to functional consequence

Drawing on the extensive biomedical information contained in the Ingenuity Knowledge Base, IPA can produce interactive maps of the complex relationships that exist between relevant genes in a dataset. Figure 4 depicts genes involved in metastasis, where HRAS has been manually included in order to look at the influence of this previously identified causal variant on the observed

expression data. A relationship exists between HRAS and MMP2, supported by 4 Findings from the Ingenuity Knowledge Base, where HRAS was shown to induce MMP2. Findings from Variant Analysis indicated a gain in function of HRAS. This information is added to IPA to predict the likelihood that the analyzed data supports the literature – that is, does the recorded HRAS variant follow the pattern of predicted activation of MMP2? The feature "Molecule Activity Predictor" (MAP) interrogates sub-networks and canonical pathways and generates hypotheses by selecting a molecule of interest, indicating up- or down- regulation, and simulating directional consequences of downstream molecules and inferred upstream activity in the network or pathway. The color-coded lines of the metastasis network shown in Figure 5 support the upstream relationship of the HRAS variant, clearly mimicking the pattern of expression change present in the analyzed dataset.

Predicting the activation state of transcriptional regulators

Whereas many studies rely on predictive binding calculations to infer impacts on transcriptional regulators, the Ingenuity Knowledge Base draws upon millions of data points from experimentally observed relationships. The Upstream Regulator Analysis feature uses the underlying experimentally derived content and identifies a likely cascade of upstream transcriptional regulators that can explain the observed gene expression changes of a given dataset.

A z-score is calculated to make statistically significant predictions of activation or inhibition for a range of transcriptional regulators, and all calculations are linked to published Findings accessible through the Ingenuity Knowledge Base. Figure 6 shows the top transcription factors predicted to be activated in the Claudin-

Summary | Functions | Canonical Pathways | Upstream Regulators | Networks | Molecules | Lists | My Pathways

ADD TO MY PATHWAY | ADD TO MY LIST | CUSTOMIZE TABLE | DISPLAY AS NETWORK | MECHANISTIC NETWORKS

Upstream Regulator	Fold Change	Molecule Type	Predicted Act...	Activation z-sc...	p-value of overlap	Target molecules in...
SRF	↑1.483	transcription regulator	Activated	3.867	1.73E-01	↑ACTA1, ↑... ...all 52
TBX2	↓-19.612	transcription regulator	Activated	3.464	3.67E-01	↑ANLN, ↓... ...all 13
Ap1		complex	Activated	3.307	4.79E-01	↓ACE, ↑B... ...all 22
REL	↓-4.076	transcription regulator	Activated	3.239	2.29E-01	↑AHR, ↑B2M ...all 23
SNAI1	↑1.041	transcription regulator	Activated	3.194	2.33E-03	↓CDH1, ↑... ...all 14
NFkB (complex)		complex	Activated	2.952	1.40E-01	↓ABCG1, ↓... ...all 94
MKL1	↓-1.728	transcription regulator	Activated	2.887	4.01E-01	↑ACTA1, ↑... ...all 17
NKX2-3	↓-6.519	transcription regulator	Activated	2.811	1.91E-01	↓ADAP1, ↑... ...all 37
POU4F1	↑51.128	transcription regulator	Activated	2.720	1.00E00	↑ACTC1, ↓... ...all 13
RELA	↑1.717	transcription regulator	Activated	2.598	1.48E-02	↑ACTN4, ↓... ...all 63
HMGB1	↑1.512	transcription regulator	Activated	2.559	1.00E00	↑ICAM1, ↑IL6 ...all 7
ISL1	↑35.413	transcription regulator	Activated	2.538	1.00E00	↑ACTC1, ↓... ...all 10
TWIST1	↑9.492	transcription regulator	Activated	2.354	7.13E-03	↓AR, ↓CBX2 ...all 16
MKL2	↓-3.887	transcription regulator	Activated	2.333	3.99E-01	↑ACTA1, ↑... ...all 14
FOSL1	↑203.391	transcription regulator	Activated	2.246	3.58E-02	↑AXL, ↓C... ...all 14
NfkB1-RelA		complex	Activated	2.227	3.71E-01	↑ICAM1, ↑IL6 ...all 5
KLF6	↑2.965	transcription regulator	Activated	2.186	1.00E00	↓CDH1, ↑ENG ...all 5
PDX1	↓-1.171	transcription regulator	Activated	2.097	9.12E-03	↓ACE2, ↑... ...all 38
GSC	↓-34.168	transcription regulator	Activated	2.000	2.33E-02	↑TGFB1, ↑... ...all 4
HOXD3	↑24.619	transcription regulator	Activated	2.000	2.59E-01	↓DSP, ↓JUP, ↓... ...all 4

Figure 6. IPA results table depicting top upstream regulators of the Claudin-low subtypes, sorted by statistical significance of predicted activation (z-scores based on experimentally derived supporting evidence within the Ingenuity Knowledge Base). The table provides information on the target molecules of each transcription factor or complex and the predicted activity. Red arrows highlight some known transcription factors of EMT, including TWIST1 and GSC.

low subtype. The red arrows highlight some key players known in the literature to be cancer markers (1).

Two genes from the activated list (SNAI1 and TWIST1) along with SPDEF (predicted to be inhibited) were chosen for inclusion in a visual representation of the transcriptional program of the Claudinlow subtype (Figure 7). SNAI1 is a transcription factor that belongs to a family of transcriptional repressors required for conversion from an EMT phenotype (1). Expression of SNAI1 promotes EMT in epithelial tumor cell lines, a process associated with the down-regulation of CDH1 (1). TWIST1 cooperates with other EMT-inducing transcription factors to suppress CDH1 and promote EMT and tumor metastasis (1). Elevated TWIST1 expression is associated with angiogenesis and tumor metastasis (8). Over-expression of TWIST1 in CD44-positive breast tumor cells promotes the generation of a breast cancer stem cell phenotype (7). SPDEF is an epithelial-specific ETS family member that is associated with the negative regulation of metastatic potential (9). SPDEF has been shown as a critical regulator of genes involved in cell motility, invasion, and adhesion (e.g., in prostate cancer cells (10)). Notably, SPDEF overexpression in MDA-MB-231 breast cancer cells (one of the Claudin-low cell lines tested here) inhibits cell growth and reduces invasion and migration (11).

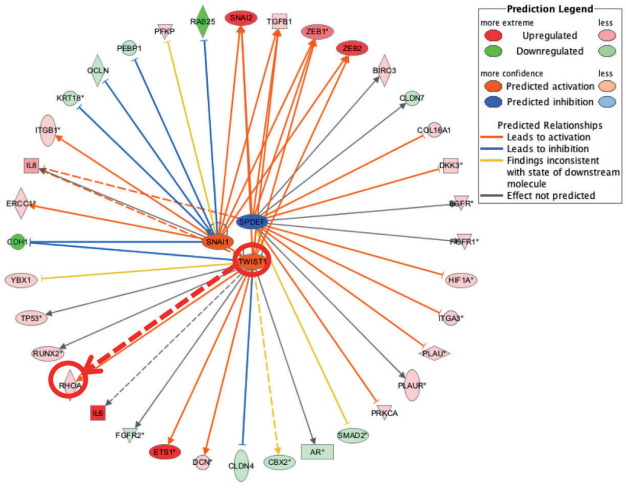


Figure 7. The transcriptional program in Claudin-low subtypes of the 3 transcription factors: SNAI1, TWIST1 and SPDEF, and their related genes. Colored arrows indicate predicted activation of surrounding molecules and the strength of the prediction in keeping with the observed downstream state of the molecules. Strong relationships exist within this network supporting the activation of EMT within Claudin-low subtypes as shown by the relevant expression of key molecules of this biological process (such as up-regulation of RHOA, SNAI2, ZEB1, ZEB2, HIF1A, etc. and down-regulation of CDH1, CLDN4, OCLN, etc.).

Connecting genes of interest

The analysis can be broadened to look at relationships of genes outside of a given network with the understanding that all variation has the potential for influence within the system. Single genes or sets of genes can be added to existing analyses to overlay and investigate additional relationships and see the predictive effects on transcription factors of interest. HRAS was added to the transcription regulators network SNAI1, TWIST1, and SPEDF (Figure 8), in addition to CD44, a cell-surface glycoprotein receptor expressed in many cancers and considered to be an important surface marker for cancer stem cells. It has been demonstrated that hyaluronan interaction with CD44 promotes SRC kinase activation, which in turn, increases TWIST1 phosphorylation, leading to the nuclear translocation of TWIST1 and its transcriptional activation (7). IPA figures will display color based upon the isoform with the biggest value and, as such, CD44 is shown to be downregulated in Figure 8.

However, many isoforms of CD44 (alternative splicing) exist, and each of these expressed isoforms may play different roles. One isoform (the variant 4, not shown) also known as CD44s, which lacks the variable exons 2–10, has been shown to accelerate EMT and breast cancer progression and is over-expressed in many other cancers (12). The CD44s variant

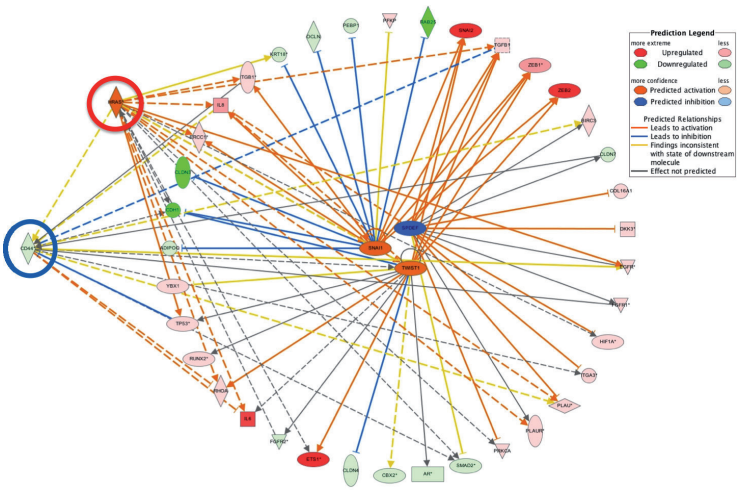


Figure 8. TFigure 8. Relational network map of the EMT-related transcription factors SNAI1,TWIST1 and SPDEF with HRAS and CD44 overlaid. Colored lines indicate the level of consistency of the relationships with the state of the downstream molecule (orange: the prediction leads to activation, blue: the prediction leads to inhibition, yellow: the prediction is inconsistent and gray: the effect is not predicted).

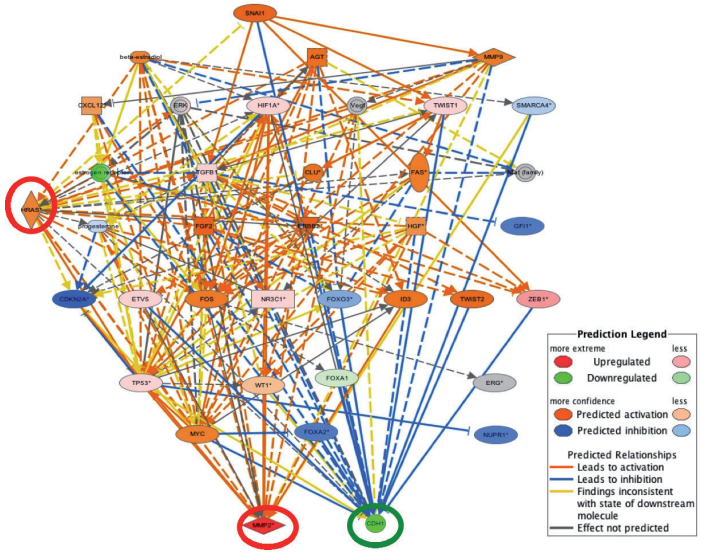


Figure 9. Mechanistic network of SNAI1 and the predicted activation and inhibition of the 35 related regulators calculated by IPA. Strong relational support of the up- and down-regulated transcription factors MMP2 and CDH1 as indicated by the solid orange (leading to activation) and blue (leading to inactivation) lines linking the transcription factors within the cascade. Notice that the "gain of function" HRAS variant integrates in the SNAI1 driven-mechanistic network and contributes to the expression of the 2 molecular markers, MMP2 and CDH1.

is present in the analyzed Claudin-low samples and is over-expressed (data not shown).

Combined effects of all regulators

Biological networks are inherently complex and IPA offers a range of options for viewing and predicting outcomes when all regulators act together. The mechanistic network driven by SNAI1 consists of 35 total regulators that generate computationally plausible directional networks responsible for the regulation of 1,011 downstream genes within the analyzed dataset. Figure 9 shows the cascade from SNAI1 through the mechanistic network of transcription factors down to the bottom row of transcriptional regulators MMP2 and CDH1. SNAI1 is the first transcription factor to be up-regulated in the EMT process, and as SNAI1 is predicted to be activated, this mechanistic network is also predicted to be activated.

Conclusion

- By pairing tools for interpreting genome (Variant Analysis) and transcriptome (IPA) data, clear differences were traced between the Claudin-low and luminal cell lines, identifying a candidate causal variant HRAS in Claudin-low subtypes that exhibited a predicted gain of function supporting an overall predicted increase of transcription factors governing EMT. In addition, a strong enrichment for pro-metastatic genes was identified across the Claudin-low samples.
- The predictive computational models of IPA identified mechanistic networks of plausible causal pathways from the key transcription factors SNAI1 and TWIST1 to MMP2 activation, known to affect EMT.
- Within any given study, generating thousands of isoforms and a multitude of possible relationships, the dual approach of Variant Analysis combined with IPA brings statistical confidence to the task of identifying probable causal influences and key isoforms.
- Each predicted relationship is linked with experimentally-derived supporting evidence from the range of peer-reviewed sources within the manually-curated Ingenuity Knowledge Base, streamlining the analysis process to quickly focus analysis efforts towards candidates with clear biological relevance.

References

1. Polyak, K. and Weinberg, R.A. (2009) Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat Rev Cancer 9, 265.
2. Prat, A. et al. (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 12, R68.
3. Ward, K.R., Zhang, K.X., Somasiri, A.M., Roskelley, C.D., and Schrader, J.W. (2004) Expression of activated M-Ras in a murine mammary epithelial cell line induces epithelial-mesenchymal transition and tumorigenesis. Oncogene 23, 1187.
4. Lovaas, J.D., Zhu, L., Chiao, C.Y., Byles, V., Faller, D.V., and Dai, Y. (2013) SIRT1 enhances matrix metalloproteinase-2 expression and tumor cell invasion in prostate cancer cells. Prostate 73, 522.
5. Lee, C.C. et al. (2012) TCF12 protein functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer. J. Biol. Chem. 287, 2798.
6. Günzel, D. and Yu, A.S. (2013) Claudins and the modulation of tight junction permeability. Physiol. Rev. 93, 525–69.
7. Bourguignon, L.Y., Wong, G., Earle, C., Krueger, K., and Spevak, C.C. (2010) Hyaluronan-CD44 interaction promotes c-Src-mediated twist signaling, microRNA-10b expression, and RhoA/RhoC up-regulation, leading to Rho-kinase-associated cytoskeleton activation and breast tumor cell invasion. J. Biol. Chem. 285, 36721.
8. Wallerand, H. et al. (2010) The epithelial-mesenchymal transition-inducing factor TWIST is an attractive target in advanced and/or metastatic bladder and prostate cancers. Urol. Oncol. 28, 473.
9. Pal, M., Koul, S., and Koul, H.K. (2013) The transcription factor sterile alpha motif (SAM) pointed domain-containing ETS transcription factor (SPDEF) is required for E-cadherin expression in prostate cancer cells. J. Biol. Chem. 288, 12222.
10. Gu, X. et al. (2007) Reduced PDEF expression increases invasion and expression of mesenchymal genes in prostate cancer cells. Cancer Res. 67, 4219.
11. Turner, D.P., Findlay, V.J., Kirven, A.D., Moussa, O., and Watson, D.K. (2008) Global gene expression analysis identifies PDEF transcriptional networks regulating cell migration during cancer progression. Mol. Biol. Cell 19, 3745.
12. Brown, R.L. et al. (2011) CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. J. Clin. Invest. 121, 1064.

Acknowledgments

The author would like to acknowledge Madeline O'Donoghue, M.S. for her expert assistance in reviewing, editing, and composing the present manuscript.

Ingenuity Variant Analysis and Ingenuity Pathway Analysis (IPA) are intended for molecular biology applications. These products are not intended for the diagnosis, prevention or treatment of a disease. For up-to-date licensing information and product-specific disclaimers, see the respective Ingenuity product site. Further information can be requested from **support@ingenuity.com** or by contacting your local account manager.

Discover insightful variant and pathway analysis at **www.ingenuity.com.**

QIAGEN Bioinformatics

EMEA
Silkeborgvej 2 · Prismet
8000 Aarhus C
Denmark
Phone: +45 7022 5509

Americas
1700 Seaport Boulevard #3
Redwood City · CA 94063
USA
Phone: +1 (617) 945 0178