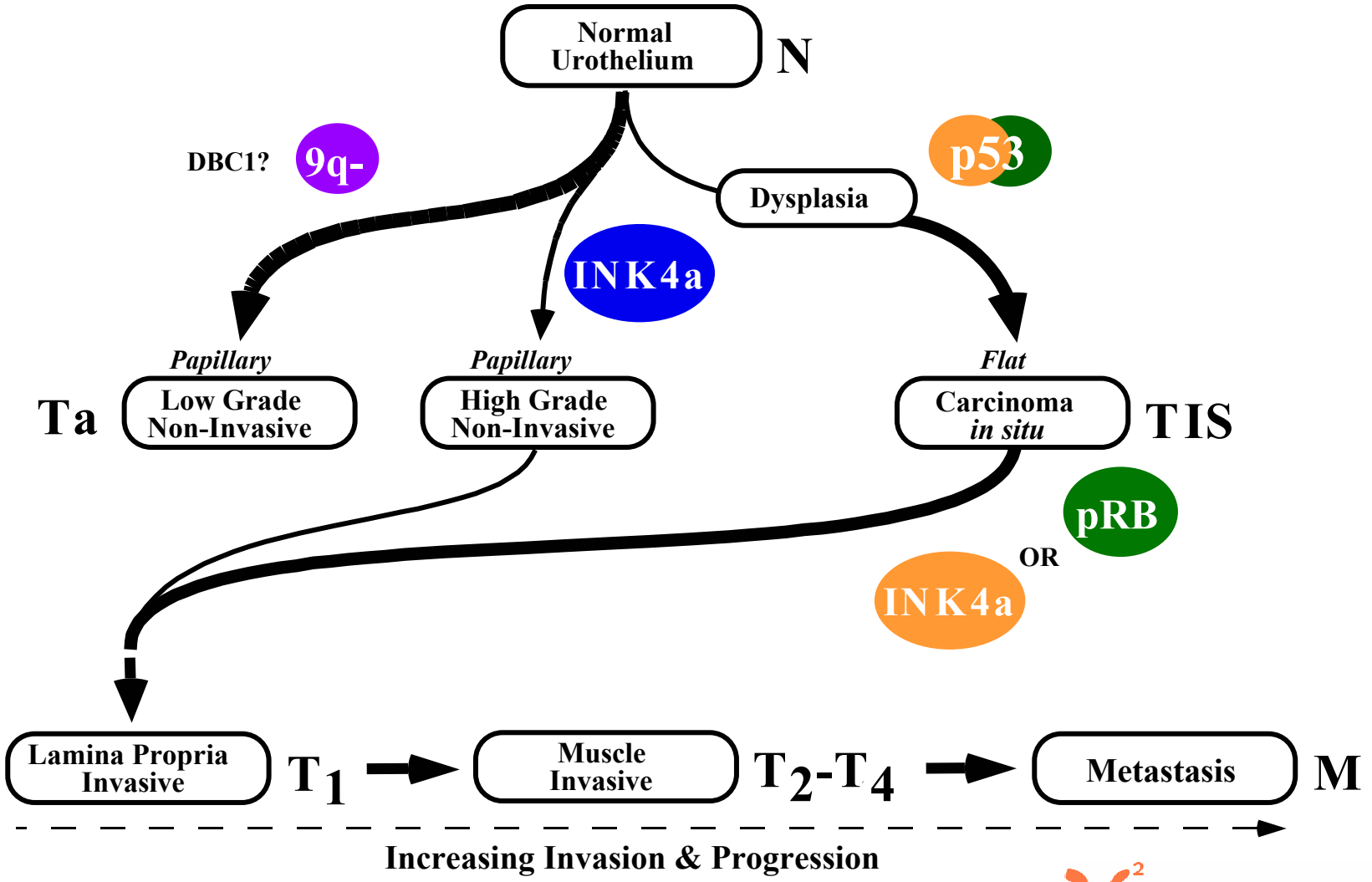
# Using GP To Develop Rules For Staging Bladder Cancer

Bill Worzel

# USC Bladder Cancer Study

- Characterizing stages of bladder cancer
  - Ta, T1-T4; Normal Samples
- RT-PCR data on selected genes
  - Gene Express StaRT-PCR
  - 70 Genes seleted by researchers at Richard Cote's lab
- Is there a molecular signature associated with each stage?

# USC: Bladder Cancer






Genetics Squared
evolutionary drug development

# Standardized RT PCR (StaRT PCR)

- Quantitative PCR based on competitive RT-PCR (Ref)
- Comparison of StaRT PCR with Real Time PCR (Ref)
  - Efficient, reproducible and less expensive
  - Good sensitivity
  - Detect variations as low as 7% in transcript quantity
  - Low consumption of cDNA sample

# Genes Profiled

## StaRT PCR: Key Pathway-specific Transcripts Quantified

**Angiogenesis**
FGF5, FGFR4, VEGF

**Anti-oxidation**
GSTM3, GSTP1, GSTT1, SOD1

**Apoptosis**
ANXA5, BAD, BCL2, CYPIA2, DAP, HSF1, KDR, NIK, PTGS2,TGFBR2, TGIF, TNFAIP1, TNFSF10, TRAF4

**Apoptosis / Cell Cycle**
BCL2L1, CDC2, CDK7, CDK8,CDKN1A, CDKN1B, CDKN2C, E2F1, E2F2, E2F4, E2F5, GAPDH1, GAPDH2, JUN, JUNB, MAD, MAX, PCNA, RB1, RBL2, TNF, TNFRF1A, TP53

**Cell Cycle**
CDKN2A, CCNA2, CCND3, CCNE1, CCNG1,CDC25C

**Growth factors**
IGF1, IGF2R, PDGFB, PDGFRL

**Invasion**
CDH3, ICAM1, MMP16, TIMP2

**Signal Transduction**
MAP2K6, MAPK12, MAP2K9, MAPK8, MYC, STAT3

**Transcription factors**
FOS, FOSL1, NFKB1, SP1

# Bladder Cancer: Results

- Rules based on known clinical stage
  - Training subset (1/3rd study set)
  - Validation subset (2/3rd study set)
- Example:
  - IF [KDR >= ((if (MAPK29 > sqr(FGFR4)) then GSTP1 else PDGFB) + MMP16)] THEN Ta
- Results:
  - 26/38 (68%) stage prediction is matches clinical staging
  - Errors may be gray areas in clinical staging (such as T1/T2 or T3/T4) - if so, then accuracy of ~83%

# Bladder Cancer: Of Interest

- Later analysis suggests that at molecular level there may be two stages: Early Stage and Late Stage Tumors
  - Different genes show up in rules for Ta, T1 and T2 when compared to T3 and T4
  - Angiogenesis lags Growth factors in Early Stage
  - Cell signalling and repressor genes are used to distinguish Late Stage
- "Normal" tissue taken beyond surgical margins shows characteristics of tumor
  - Samples are classed the same as resected tumors

# Other Applications

- Toxicogenomics
  - Pre-clinical study of toxicity based on gene chip analysis
- Cheminformatics
  - Correlating structure-activity relationships  from high-throughput screening (HTS) data
- Clinical Drug Response
  - Multiple myeloma study with the Van Andel Research Institute of Grand Rapids, MI
  - Study on effectiveness of immunosuppresant after stem-cell transplant with Fred Hutchison Cancer Research Center
  - From baseline data, can we predict the best course of treatment for a patient? Stay tuned…

# Comments on the Industry

- Some general rules:
  - Danger of overfitting data is high
  - Small number of samples, high dimensionality of data - “The curse of dimensionality!”
  - Must use validation techniques (eg, N-fold cross validation)
  - If possible, reserve samples as a validation set
  - Prospective proof of generality needed
    - Results change from lab-to-lab
    - Results change from chip-to-chip
- You must satisfy the statisticians or you get no where!