Gene expression

- What is gene expression?
- Methods for measuring a single gene.
  - Northern Blots
  - Reporter genes
  - Quantitative RT-PCR
- Operons, regulons, and stimulons.
- DNA microarrays.
  - Expression profiling
  - Identifying protein binding sites.
  - Comparing gene content of different strains.

What is gene expression?

- The amount of RNA produced from a gene.
- Level of RNA produced from a gene is controlled by:
  - Transcription
  - Degradation
- Transcriptome - Expressed transcripts in a cell under defined experimental conditions.
  - mRNA(5-10% of total RNA).
  - rRNA, tRNA - make up most of total RNA
  - scRNA (protein secretion), tmRNA (rescue stalled ribosomes).

Analysis of gene expression at the single gene level.

- Northern Blots
  - Measure RNA levels by hybridization of a labeled probe to total RNA.
- Reporter Genes
  - Use of an enzyme to measure the amount of transcription from a promoter.
- Quantitative RT-PCR.

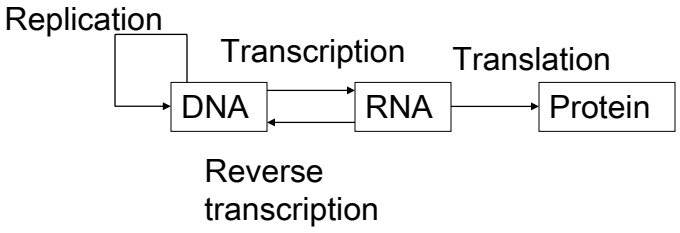
Regulons and Stimulons

- Operon - group of genes co-expressed on a single transcript.
  - One location of the genome
- Regulon - genes that are regulated by a single transcription factor.
  - Genes and operons throughout the genome
- Stimulon - collection of genes that are regulated in response to environmental changes.
  - Can be multiple regulons affected at once.
- Regulatory network - alternative term for regulon.

Assaying the regulation of 1000s of genes in a single experiment

- DNA microarrays
  - DNA molecules printed at high density used to determine the level of RNA or DNA in a sample.
  - Can be thought of a “reverse Northern blots”

Central Dogma



DNA ATATCGGCATCAGTCGATCGATCATCGATCGAT

↓

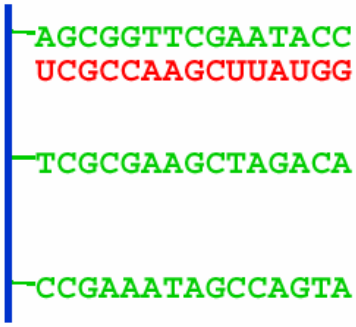
mRNA UAUAGCCGUAGUCAGCUAGCUAGUAGCUAGCUA

↓

cDNA ATATCGGCATCAGTCGATCGATCATCGATCGAT

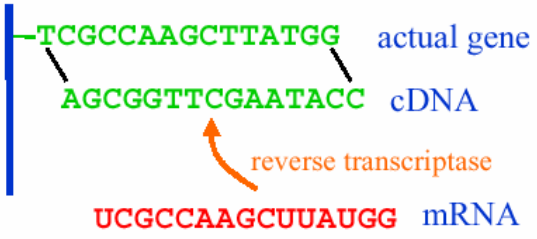
Complementary Hybridization

- due to Watson-Crick base pairing, an mRNA molecule will hybridize to a complementary DNA molecule



Complementary Hybridization

- the way it's usually done
  - put the actual gene sequence on array
  - convert mRNA to cDNA using *reverse transcriptase*



DNA Microarrays -Introduction

- Spotted DNA arrays (glass slides)
  - Competitive binding of samples
  - Fluorescent detection - Cy3 and Cy5
  - Small sample sizes (10-30µl).
  - PCR or cDNA arrays
  - Long oligonucleotide arrays
    - Better specificity, cheaper, easier to work with.
- Short oligonucleotide arrays
  - ex. Affymetrix
- DNA spotted onto nylon membranes (macroarrays)

What Is a Microarray

- Different Approaches

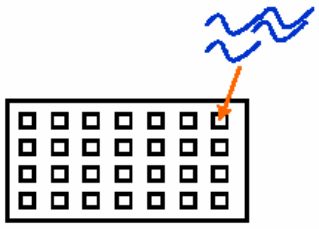
	Stanford/ Pat Brown	Affymetrix
How DNA sequences are laid down	Spotting	Photolithography
Length of DNA sequences	cDNA(Complete sequences)	Oligonucleotides

Stanford Approach

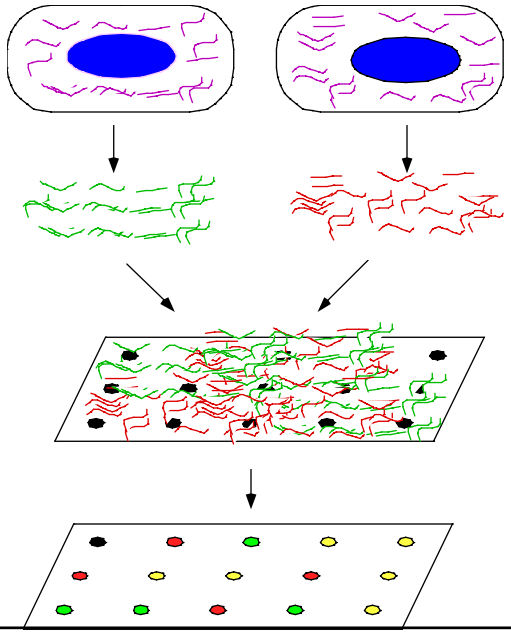
- Use robot to spot glass slides
- Able to measure qualitatively relative expression levels of genes
  - Differential expression by use of simultaneous, two-color fluorescence hybridisation
- Cheaper with DIY ($60,000)
- Also called home-made system

Spotted Arrays

- robot puts little spots of DNA on glass slides
  - each spot is DNA analog of one of the mRNAs we want to measure



Microarray experimental overview



Grow cells

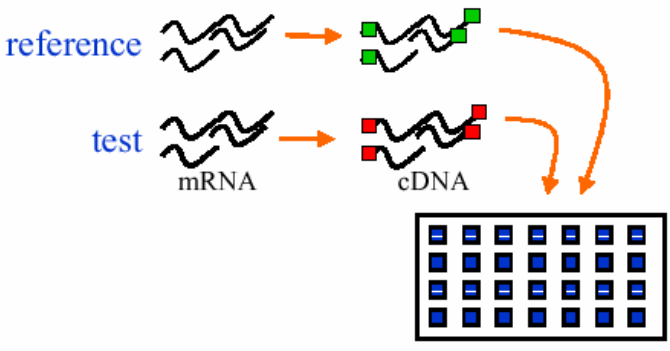
Isolate RNA
Make labeled cDNA

Mix and hybridize

Scan slide
Analyze data

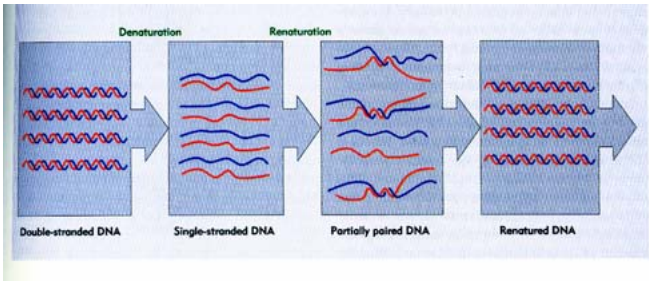
Spotted Arrays

- two samples (reference and test) of mRNA are converted to cDNA, labeled with fluor dyes and allowed to hybridize to array



Hybridization: basic concept

The ability of two strands to hybridize is dependent on their complementarity.
More complementarity=better hybridization



Labeling RNA or DNA with Cy3 or Cy5.

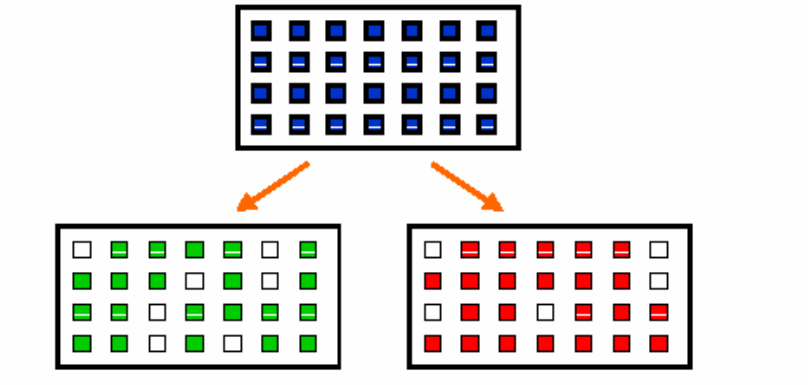
- Cy3 and Cy5 - most often used fluorescent molecules used to label samples for microarray analysis.
  - Absorb light at one wavelength and emit at another.
  - Emission and Excitation spectra do not overlap significantly.
  - In arrays Cy3 and Cy5 are usually false colored green (Cy3) and red (Cy5) for ease of visualization.

More labeling

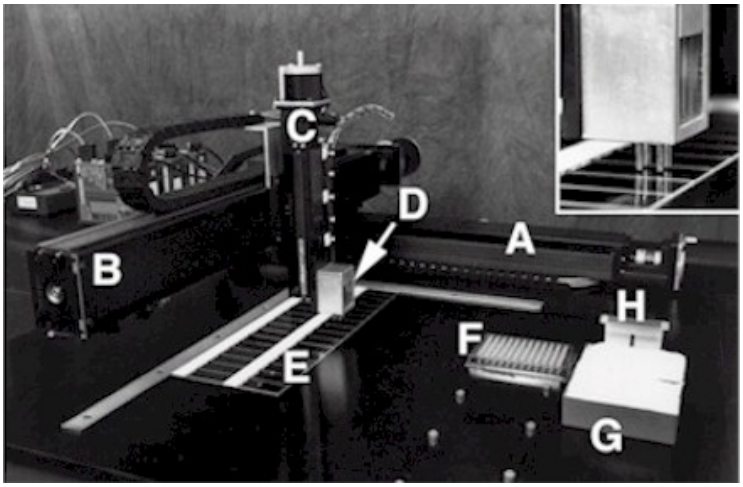
- Direct incorporation - incorporates Cy3-or Cy5-dNTP directly into cDNA
  - RNA to cDNA - reverse transcriptase
  - DNA to DNA - DNA polymerase
  - Big problem - Cy3 and Cy5 are not incorporated with same efficiency.
- Indirect incorporation - preferred method.
  - Incorporate an aminoallyl-dUTP molecule during reverse transcription of RNA to cDNA.
  - Chemically couple Cy3 or Cy5 dye after cDNA is made.
  - Coupling is efficient with both dyes.

Spotted Arrays

- Lasers applied to the arrays yield an emission for each fluorescent dye



Arrayer



Scanner

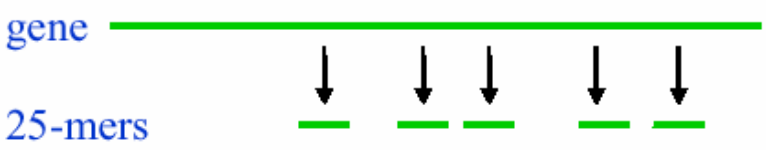
- GenPix 4000

Oligonucleotide Arrays

- a.k.a. “gene chips”
- instead of putting entire genes on an array, put sets of DNA 25-mers (oligonucleotides)
- produced using a photolithography process similar to that used to make semiconductor chips
- mRNA samples are processed separately instead of in pairs

Oligonucleotide arrays

- given a gene to be measured, select 20 25-mers for that gene



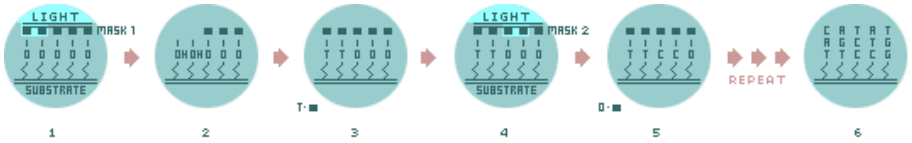
- selection criteria
  - specificity
  - hybridization properties
  - ease of manufacturing

Oligonucleotide arrays

- each of these probes is put on the chip
- additionally a slight variant (that differs only at the 13th base) of each is put next to it
  - this helps factor out false hybridizations
- the measurement for a gene is derived from these 40 separate measurements

Affymetrix

- Probe Array (Photolithography)
  - Synthesis of probe



GENECHIP® PROBE ARRAY SYNTHESIS PROCESS

Affymetrix vs. glass slide based arrays

- Affymetrix
- Short oligonucleotides
- Many oligos per gene
- Single sample hybridized to chip

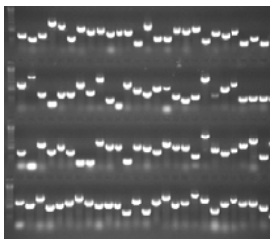
- Glass slide
- Long oligonucleotides or PCR products
- A single oligo or PCR product per gene
- Two samples hybridized to chip

Applications of DNA microarrays

- Expression profiling
  - Determining the relative levels of RNA in two or more samples.
- DNA/DNA hybridizations
  - Investigate gene content between different strains
  - Determine gene dosage
  - 16S arrays - microbial communities (being developed).
- Identification of protein binding sites
  - ChIP-Chip. Immunoprecipitation of protein/DNA complexes. Assaying those interactions with microarrays.

B. subtilis DNA microarrays

- PCR generated microarrays using custom primers (Sigma-Genosys).
- Each PCR product represents a single gene.
- 4074 genes of 4101 on the array.
- Printed on Corning CMT-GAPS slides.
- 4 *E. coli* controls, each represented 15-20 times on the array.



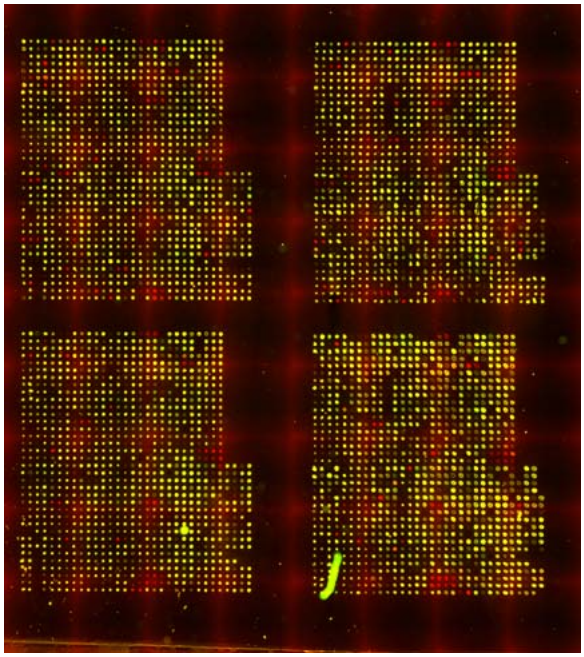
How a DNA microarray works

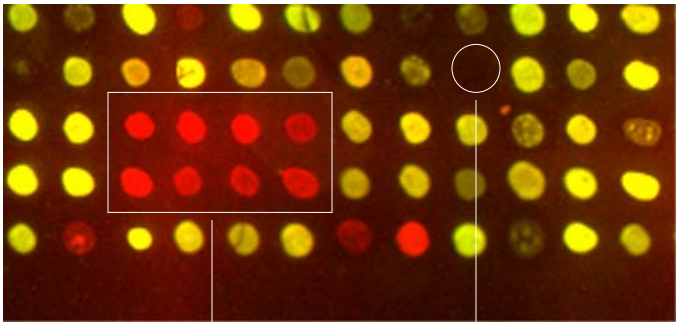
- Comparing the genome content of two *B. subtilis* strains.
- The two strains differ only by the fact that JH642 is lysogenized with the bacteriophage SPβ.
- JH642 vs PY79 genomic DNA hybridization.
  - PY79 does not contain SPβ.
  - SPβ spots will be red.

JH642

PY79

Array size = 16mm x16mm
Spot size = 150μM





SPβ genes *E. coli* control

JH642

PY79

Bacterial DNA microarrays

- Small genome size
- Fully sequenced genomes, well annotated
- Ease of producing biological replicates
- Genetics

Applications of DNA microarrays

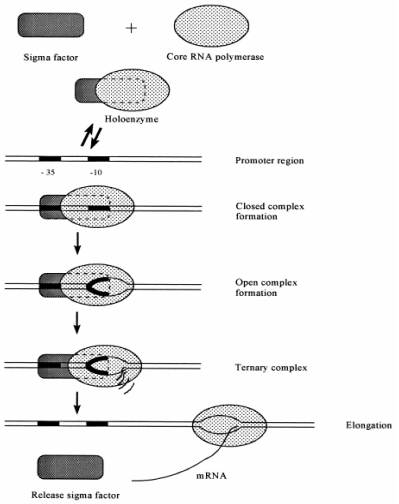
- Monitor gene expression
  - Study regulatory networks
  - Drug discovery - mechanism of action
  - Diagnostics - tumor diagnosis
  - etc.
- Genomic DNA hybridizations
  - Explore microbial diversity
  - Whole genome comparisons
  - Diagnostics - tumor diagnosis
- ?

Characterization of the stationary phase sigma factor regulon (σ^H) in *Bacillus subtilis*

- Patrick Eichenberger, Eduardo Gonzalez-Pastor, and Richard Losick - Harvard University.
- Robert A. Britton and Alan D. Grossman - Massachusetts Institute of Technology.

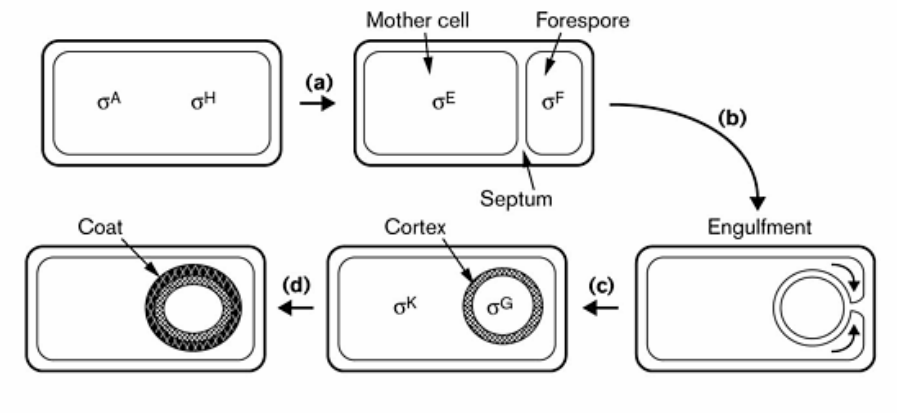
What is a sigma factor?

- Directs RNA polymerase to promoter sequences
- Bacteria use many sigma factors to turn on regulatory networks at different times.
  - Sporulation
  - Stress responses
  - Virulence



Wosten, 1998

Alternative sigma factors in *B. subtilis* sporulation



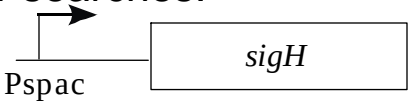
Kroos and Yu, 2000

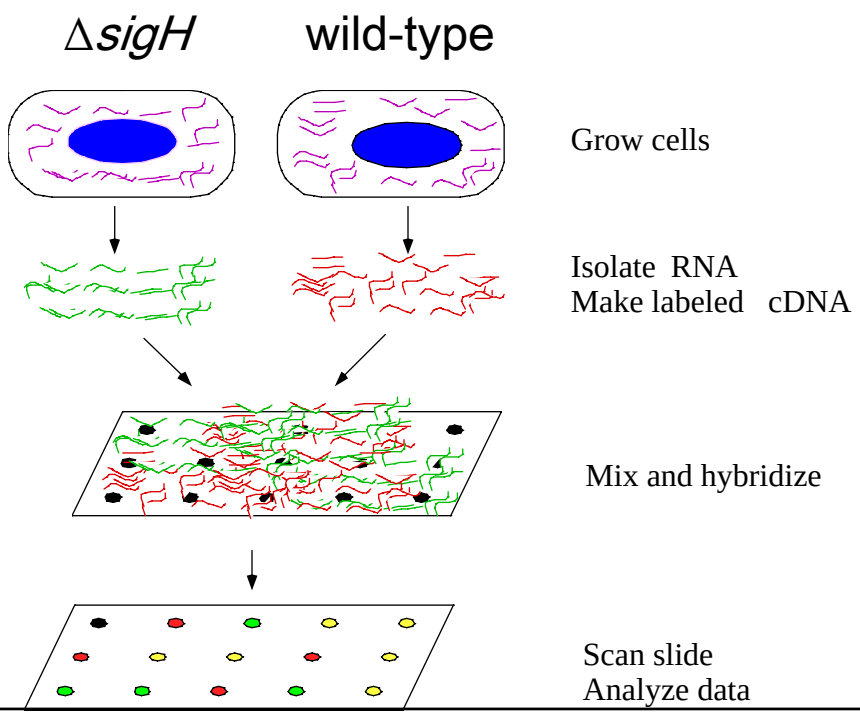
The stationary phase sigma factor: σ^H

- most active at the transition from exponential growth to stationary phase
- mutants are blocked at stage 0 of sporulation
- known targets involved in:
  - phosphorelay (*kinA*, *spo0F*)
  - sporulation (*sigF*, *spoVG*)
  - cell division (*ftsAZ*)
  - cell wall (*dacC*)
  - general metabolism (*citG*)
  - phosphatase inhibitors (*phr* peptides)

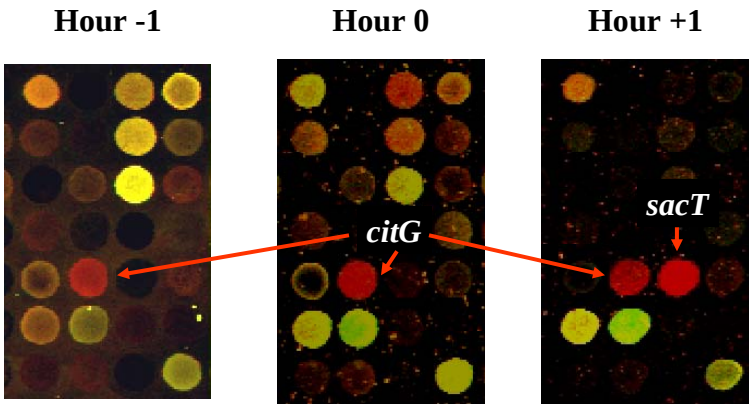
Experimental approach

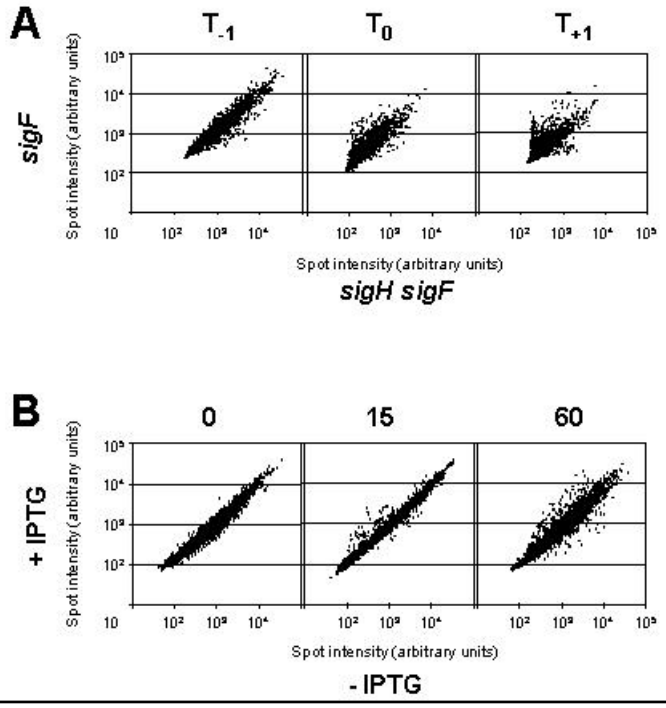
- Compare expression profiles of wt and *ΔsigH* mutant at times when *sigH* is active.
- Artificially induce the expression of *sigH* during exponential growth.
  - When Sigma-H is normally not active.
  - Might miss genes that depend additional factors other than Sigma-H.
- Identify potential promoters using computer searches.





wild type (Cy5) vs. *sigH* mutant (Cy3)





Identifying differentially expressed genes

- Many different methods
- Arbitrary assignment of fold change is not a valid approach
- Statistical representation of the data
  - Iterative outlier analysis
  - SAM (significance analysis of microarrays)

Data from a microarray are expressed as ratios

- Cy3/Cy5 or Cy5/Cy3
- Measuring differences in two samples, not absolute expression levels
- Ratios are often log2 transformed before analysis

Genes whose transcription is influenced by σ^H

- 433 genes were altered when comparing wt vs. Δ*sigH*.
- 160 genes were altered when *sigH* overexpressed.
- Which genes are directly regulated by Sigma-H?

Identifying *sigH* promoters

- Two bioinformatics approaches
  - Hidden Markov Model database (P. Fawcett)
    - HMMER 2.2 (hmm.wustl.edu)
  - Pattern searches (SubtiList)
- Identify 100s of potential promoters

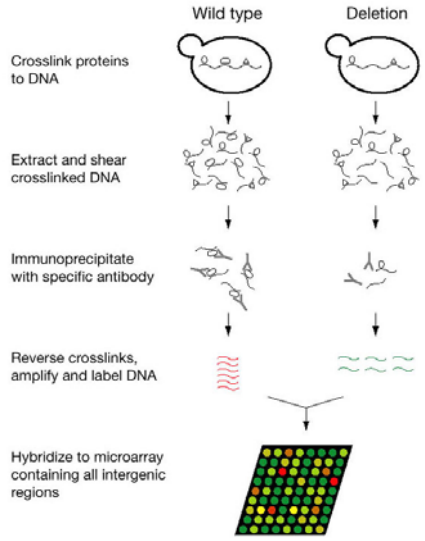
Correlate potential *sigH* promoters with genes identified with microarray data

- Genes positively regulated by Sigma-H in a microarray experiment that have a putative promoter within 500bp of the gene.

Directly controlled *sigH* genes

- 26 new *sigH* promoters controlling 54 genes
- Genes involved in key processes associated with the transition to stationary phase
  - generation of new food sources (ie. proteases)
  - transport of nutrients
  - cell wall metabolism
  - cyctochrome biogenesis
- Correctly identified nearly all known *sigH* promoters
- Complete *sigH* regulon:
  - 49 promoters controlling 87 genes.

- Identification of DNA regions bound by proteins.



Iyer et al. 2001 Nature, 409:533-538