

Rice transcriptome:  
Rice full-length cDNA clones  
and  
Rice oligoarray for gene  
expression analysis

Shoshi Kikuchi

Natl Inst Agrobiological Sciences

[skikuchi@nias.affrc.go.jp](mailto:skikuchi@nias.affrc.go.jp)



## News@Agilent

### Agilent Technologies commercializes first 60-mer microarray for rice

**Microarray designed and validated by Japan's National Institute of Agrobiological Sciences**

PALO ALTO, Calif., Nov. 6, 2003

Agilent Technologies Inc. (NYSE: A), in collaboration with Japan's National Institute of Agrobiological Sciences (NIAS), today commercialized the first 60-mer oligonucleotide microarray for the study of rice, a staple food for half the world's population. Researchers can use the Agilent Rice Oligo Microarray Kit to measure the activity of genes in rice and related cereal plants, helping them identify varieties with greater tolerance to drought, salt, cold climate or pests for planting in less arable lands.

The microarray includes genetic probes for more than 21,000 genes from the genome of *Oryza sativa* *L. ssp japonica* (cultivar Nipponbare), a strain of rice that is mainly cultivated in Japan. This is believed to be approximately 50 percent of the total rice genome, currently estimated at 40,000 to 50,000 genes. Agilent manufactures the microarrays using ink-jet-based technology, which prints DNA *in situ* onto 1" x 3" glass slides to a length of 60 oligonucleotides. The 60-mer gene probes provide five to eight times greater sensitivity than 25-mer probes.

"This microarray is based on actual, biologically-expressed sequences (cDNA), not sequences predicted to be genes by computer," said Dr. Shoshi Kikuchi, head of the Laboratory of Gene Expression at NIAS and leader of the Rice Microarray Project, a part of Japan's National Rice Genome Project. "After several validation experiments with Agilent custom rice oligo microarrays, we found that the signals were very clear and the reproducibility was very good. We believe the introduction of this microarray system to the scientific community will accelerate the functional characterization of genes in rice and related cereal plants."

The NIAS is an independent administrative research institute in Tsukuba, Japan, that maintains the world's most complete cDNA library of rice genes. Their full-length rice cDNA sequences, upon which this oligo microarray is based, were collected and completely sequenced as part of the Rice Full-length cDNA Project. The sequences and annotations of these rice cDNAs, which correspond to approximately 22,000 unique expressed gene sequences, were recently published(1) by Shoshi Kikuchi and coworkers. NIAS will provide cDNA clones to academic researchers upon request.

"Agilent's rice oligo microarray will complement the rice genome sequencing effort particularly in clarifying the function of the more than 50,000 genes predicted to exist in rice," said Dr. Takuji Sasaki, head of Japan's Rice Genome Research Program (part of the National Rice Genome Project) and the International Rice Genome Sequencing Project.

The first draft of the rice genome sequence was published in 2002, making rice the second plant organism to have its genome sequenced. The first was *Arabidopsis thaliana*, a member of the mustard family commonly studied by researchers as a model plant system. Like *Arabidopsis*, rice is a model organism for plant researchers. Among cereal plants, it is believed to have one of the smallest

#### Press Releases

- ▶ [Communications](#)
- ▶ [Corporate](#)
- ▶ [Electronics](#)
- ▶ [Life Sciences/ Chemical](#)

#### ▶ [Archives](#)

#### Search Agilent



#### Quick Links

Jump to page



On 7th November, 2003 JST

# 22K oligo array system-based on the full-length cDNA sequence information

- **Custom array**  
(Oct. 2002 ~ Dec. 2002)  
Designed from 29,100 clones  
21,938 probes (18,324TUs) out of 28,850 designed oligomer  
(Jan. 2003 ~ Dec. 2003)  
Validation experiments were done.
- **Catalog array (G4138A)**  
Considering the results of the validation experiments (Jan. 2002~ July 2003) and new mapping results,  
21,495 probes (19,109TUs) out of 28,850 designed oligomer were chosen.  
(Nov. 2003)  
Announcement of Catalog array

Information of the genome sequence of japonica rice was not used.

# Why Agilent?

- Custom array production system (cheaper than other company)
- 60mer oligo representing one transcript
- Reproducible and reliable array production system
- Very easy to obtain expression data
- Good and reproducible expression data

# Our rice microarrays along with the rice genome project

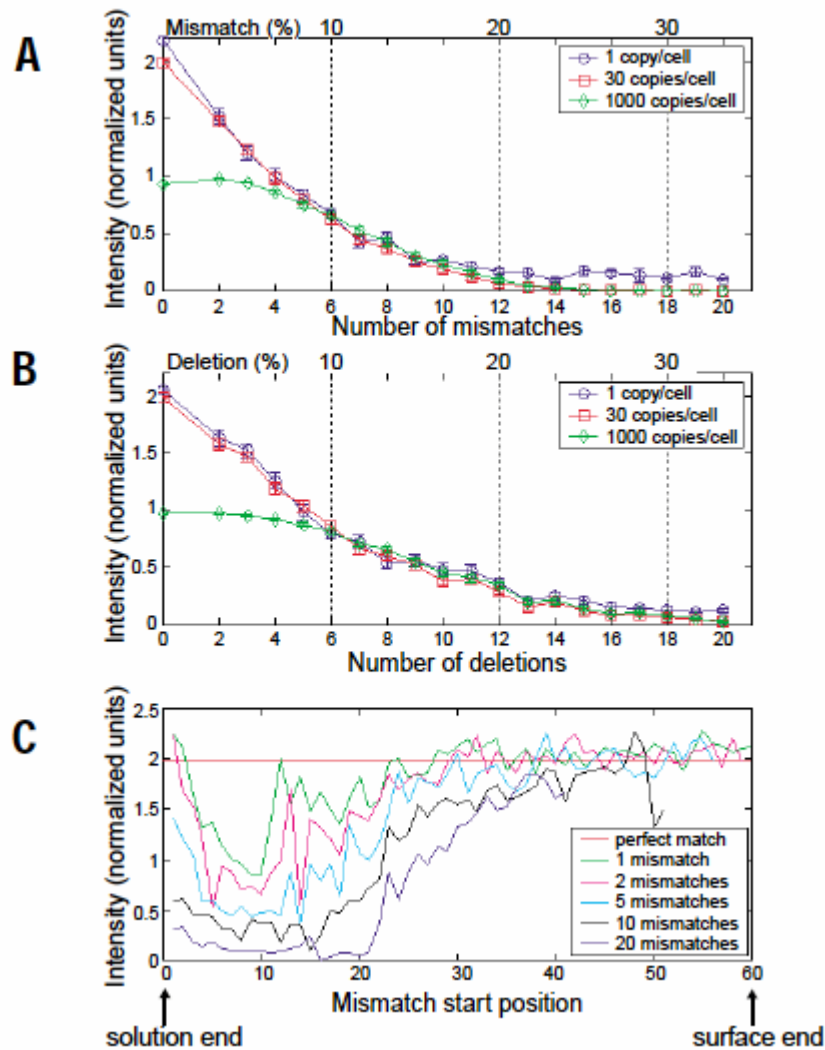
Term	Project	Products
1991-1997	Phase I of the rice genome project	RFLP mapping Physical map of rice genomic DNA <b>Collection of EST clones (40K~60K)</b> <b>Number of unique clones are ~11,000)</b>
1999-2004	Phase II of the rice genome project	Complete genome sequence by IRGSP Dec. 2002 Completion of main parts (phase II) Dec. 2004 Completion (phase III)
1999-2003	Rice microarray project	<b>1265 and 8987 cDNA-based microarray systems</b> RMOS (Rice microarray opening site) RED (Rice Expression Database)
2000-2003	Rice-full length cDNA project	<b>Collection, mapping, annotation of 32K rice full-length cDNA clones</b> KOME (Knowledge-based Oryza Molecular biological Encyclopedia ) <b>Commercialization of 22K rice oligo array</b>

# Comparison of two microarray systems

	cDNA-based array	Oligo array
Number of probes	1265+8987	21,495 (catalog array G4138A)
Installation	1999~2003	2003~
Technically supported by	MTAP(Microarray Technology Access Program) Amersham Pharmacia Biotech UK Limited Molecular Dynamics, Inc Amersham Pharmacia Biotech Japan KK	Agilent Technologies
Sequence information of the probes	Double stranded DNA Full-insert of the plasmid cDNA was stumped on the glass. Single pass sequences from 5' and 3' ends are known.	Single stranded 60mer oligomer 60mer sequences were designed through the bioinformatic analysis. Array user will be informed on the 60mer sequences. Available for the strand-specific transcriptome analysis such as the antisense transcripts

# Comparison of two microarray systems

	cDNA-based array	Oligo array
Scope of application on demands of researchers	Wide	No
Processes to carry out the microarray experiments	So many	Labeling and hybridization
To get the reproducible results	Not so easy	Easy
Cross hybridization between the members of family gene. (MA data vs Northern, RT-PCR data)	High (depends on the sequences of probes)	Low
Application to other Graminae sp.	Grass, Wheat, Barley tested OK	Presently, other than Nipponbare <i>indica</i> tested OK
Quantification of the spot data	Hard	Easy (if you use an Agilent Scanner and the software)
Data mining	Not well facilitated	Combination with the data of KOME site
Price	Considering whole system, Very expensive	JPY: 313,000/ five plates kit



**Figure 2.** Dependence of hybridization intensity on number and location of mismatches and deletions. (A) Hybridization intensities are plotted as a function of the number of randomly placed single-nucleotide mismatches in the 60-mer sequence. Intensities are normalized to sample abundance. (B) Mean hybridization intensities are plotted as a function of the number of randomly placed single-nucleotide deletions in the 60-mer sequence. Intensities are normalized to sample abundance. (C) Mean hybridization intensities at 30 copies/cell equivalent are plotted as a function of the starting position for the indicated number of contiguous nucleotide mismatches in the 60-mer sequence.

**Specificity of hybridization to 60-mers.** We first examined the effects of different numbers of randomly placed single-base mismatches and deletions (ranging from 2 to 20 per oligonucleotide) on hybridization yield at three different sample concentrations (Fig. 2A, B). Eighteen or more randomly placed mismatches per 60-mer reduced hybridization to background levels, even with samples at 1,000 copies per cell equivalent in human cells. This suggests that, on average, maximal discrimination is achieved by choosing oligonucleotides that differ from all other sequences within the genome by 18 or more out of 60 bases.

# Examples of Agilent 60mer oligoarray system for cross species gene expression analysis

- Arabidopsis 1 for Brassica nupas

Reported by Hadfield et al (JIC)

- Originally Arabidopsis 1 is designed from Ecotype (Col) is also available for *Ler*, *WS* etc.
- Rice ver 1. for indica rice in our collaborative work with IRRI

# Microarray experiment

## 1. RNA preparation, polyA RNA purification



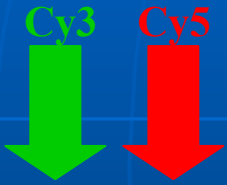
- Quality check (Agilent BioAnalyzer)
- Concentration check (NanoDrop)

• RNeasy kit (QIAGEN)

• Oligotex<sup>TM</sup>-dT30 <super>  
mRNA Purification kit (TAKARA)

## 2. Linear amplification and labelling

• Fluorescent Linear Amplification kit (Agilent)



- Quality check (Agilent BioAnalyzer)
- Concentration check (NanoDrop)

## 3. Hybridization (60C, 17hrs) and washing

• 22.5K *In Situ* Hybridization kit (Agilent)

Four slide glass / experiment

- **Control Cy3** vs **Control Cy5** Self experiment
- **Target Cy3** vs **Target Cy5**
- **Control Cy3** vs **Target Cy5** Dye-swap
- **Control Cy5** vs **Target Cy3** experiment

## 5. Data analysis

- Feature Extraction (Agilent)
- Genespring ver.6.0 (silicongenetics)

## 4. Scanning



# Reproducibility of “self” experiments

X-axis : Signal of Cy3

Y-axis : Signal of Cy5

(normalized data, Flagged spots\* were removed)

Array Upper - No.1(011101)

Self S vs. S

R=0.99

- Spot>2.0 fold change:  
0.2% (45 out of 20,018)

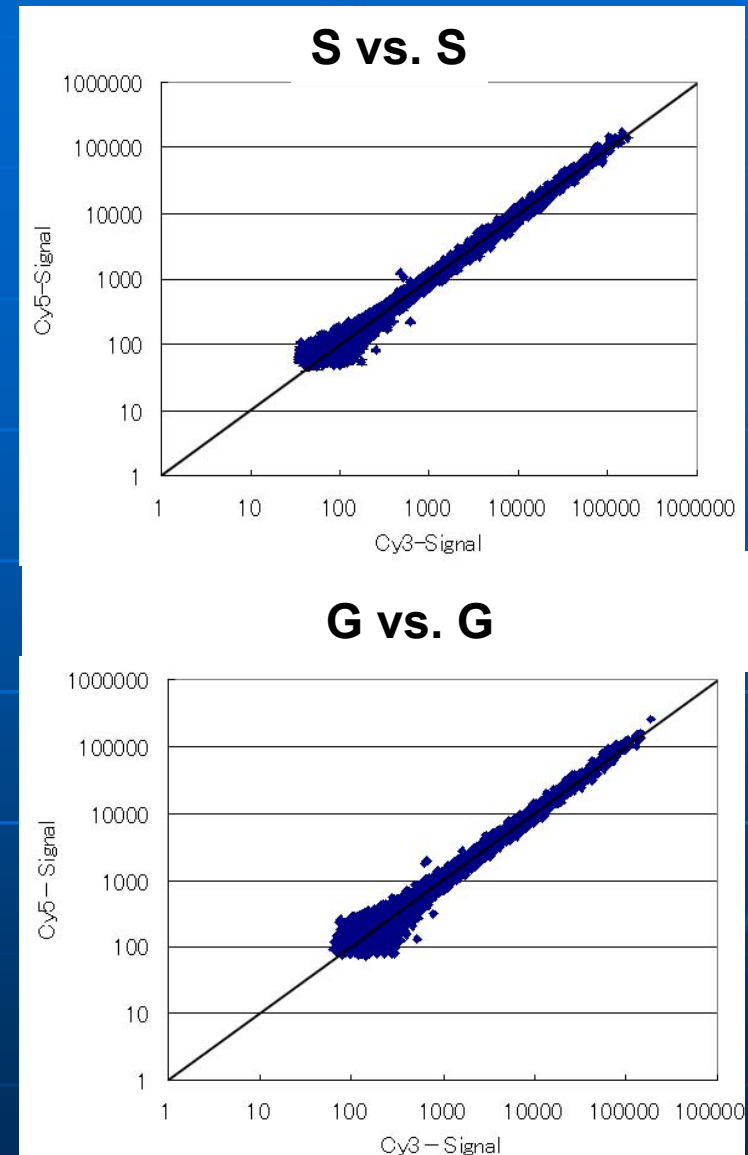
Array Bottom - No.2 (011100)

Self G vs. G

R=0.99

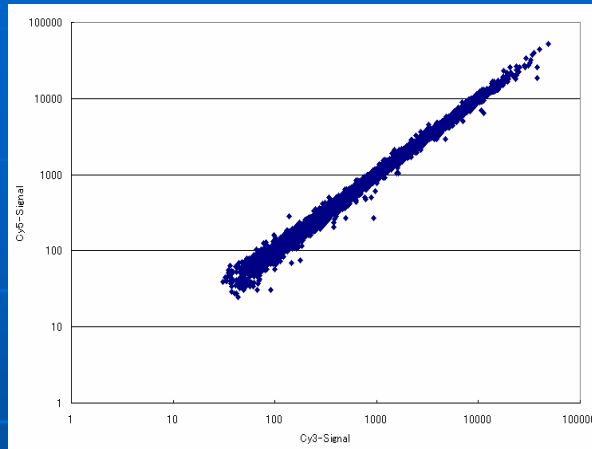
- Spot>2.0 fold change:  
0.78% (151 out of 19,349)

S: Shoot G: Germinating seeds



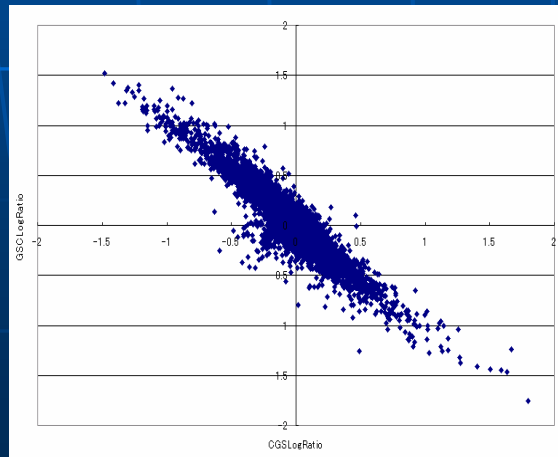
# Results of Catalog microarray (21,495 spots)

## Self-experiment **S** vs. **S**



- Self  
X-axis-Cy3-signal  
y-axis-Cy5-signal  
R=0.99
- Effective spots  
91.66% (19,702 /21,495)
  - Available spots  
99.96% (19,695/19,702)
  - Spot>2.0 fold change:  
0.04% (7 / 19,702)

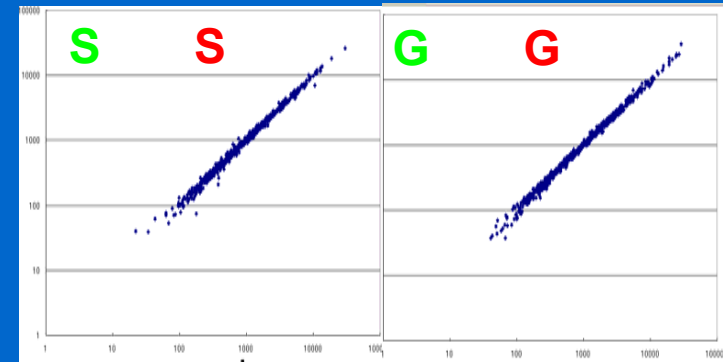
## Color-swap **S/G** Ratio vs. **G/S** Ratio



R=0.92 (18,902 spots/21,495)

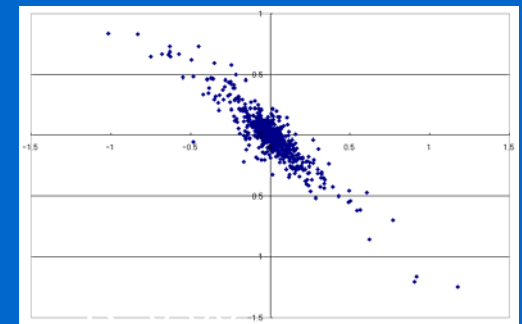
Expression data of newly added 785 spots

## Self experiment



Signal was widely distributed from low to high.

## Color-swap **S/G** Ratio V.S. **G/S** Ratio



Good reproducibility

# Samples for analysis

## In house

- Several life stages of rice
- Several environmental stress treatments to seedlings
- Hormonal treatments (ABA,GA) to calli

## Collaboration with other research groups

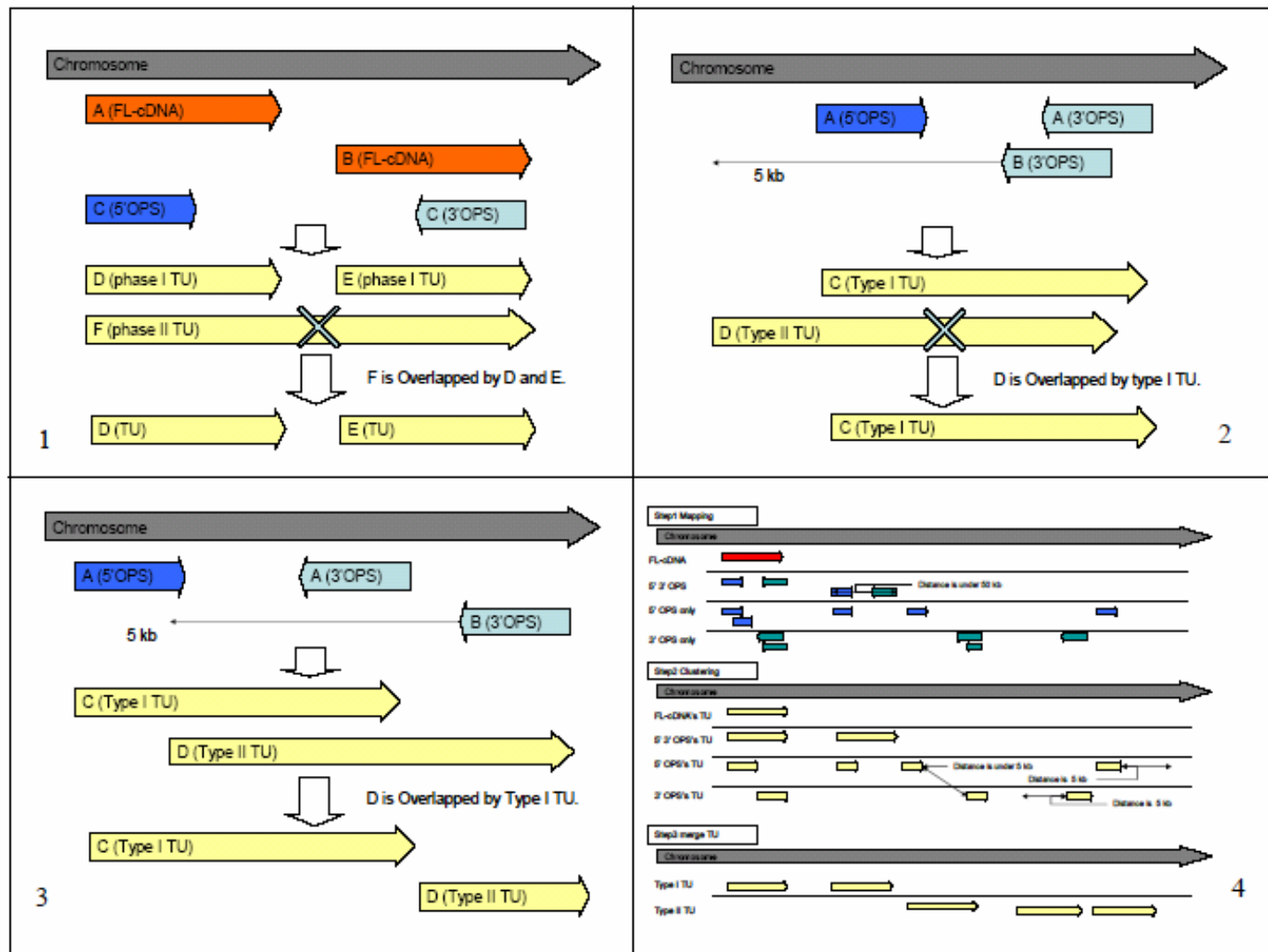
(IRRI, Korea, China)

- Viral infection
- Drought stress treatments
- Fungi infection
- Genomic copy number analysis
- Gene expression under mutation
- Nitrogen nutrition condition

# Future prospects

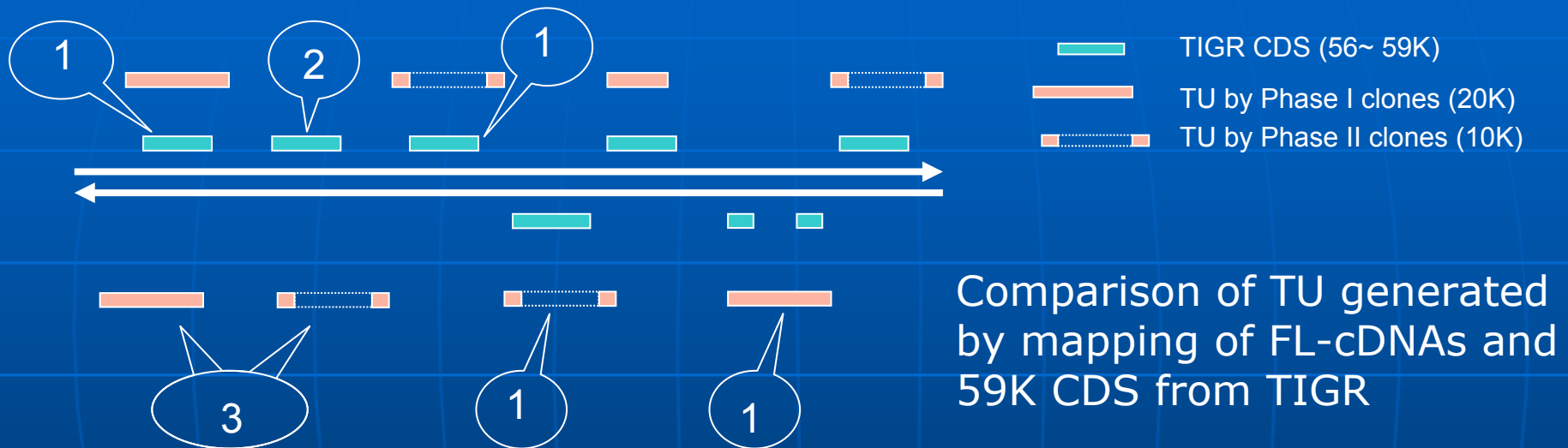
- Improvement of the rice oligo array system
  - > Add new spots from the phase II collection of full-length cDNA clones (580K)
  - > User friendly data analysis system
  - > Much more information (annotation) of the genes on the array 44K array format is available now.
  - > Open the gene expression database

# Mapping of 580K EST to rice genome sequence



# Updating the array to 44K format

based on 580,000 EST sequences from full-length cDNA clones



1 AE (Annotated Expressed) genes: 25,249 CDS or 24,510 TU

2 ANE (Annotated Non Expressed) genes: 32,286 CDS

3 NAE (Non Annotated Expressed) genes: 5,132 TU

# Expression analysis of ANE genes

- Specific oligomer sequences from ANE genes will be chosen
- Printed on 44K format array from Agilent
- Expression data will be obtained or not?

# Two types of arrays from one set of oligo-design

- 3' UTR oligo for species-specific array
- 5' coding region oligo for universal array