SOLiD™ Data – 2 Base Encoding
What is two base encoding?

• Rather than a probe, reading out the single base present at the 5th position, a two base encoded probe tells us information about the 4th and 5th bases which needs further information to resolve the base call.

• In order to do this we use the concept of color space.
Color Space - Capillary electrophoresis

1. Collect color image
2. Identify peaks
3. Identify peak colors
4. Convert to Base calls
Color Space – SOLiD (Dual Base encoding)

Collect color image → Identify Beads → Identify Bead color

Record colors for each bead over consecutive cycles
2 Base Pair Encoding
Using 4 Dyes

On our probes the 1st base encoded is position 4
the 2nd base encoded is position 5
Ball and Stick Model

A • A • C • A • A • G • C • C • T • C
Consequences of 2 Base Pair Encoding

- Detecting a single color does not indicate a base
- Each reading contains information from two bases
- To decode the bases you must know one of the bases in the sequence
Example of decoding (i)

Cannot determine any of the bases
Example of decoding (ii)

If know first base is an A then immediately it decodes 2\textsuperscript{nd} base. This must be an A as Blue translates 2\textsuperscript{nd} base A if first base A.
Summary of decoding

AACCAAGCTCGC
AACAAGCCTC
CAACGTCC
AA
CCGGTT
ACCAGTTG
ACCAGTTG
AACCGGTT
AGCTGATC
AGCTGATC
AGCTGATC
ATCGGCTA

AACACAAGCCTC
Advantages of 2 base pair encoding

- Double base interrogation eases the discrimination between system *errors* and *true* polymorphism
Advantages of 2 base pair encoding
Real SNP

Two color changes represent only a single mismatch to reference sequence (SNP)
Advantages of 2 base pair encoding

Miscall

A C G G T C G T C G T G T G C G T

reference

expected

A C G G T C G T C G T C G T C T A C A C A C A T A C

observed

Single color change, represents sequencing error.
But there's more…
only certain transitions are allowed for a real SNP

Consider a triplet of bases, they define 2 colors. There are only 3 possibilities for a change in the middle base, hence only 3 possibilities for the 2 colors to change to. Any of the other 6 possibilities for a 2-color change are not allowed and most probably represent measurement errors. (There are only 9 possibilities where both colors have changed)
The **only** allowed color changes

**If two colors present (eg B,R)**
- Reverse the colors (eg R, B)
- Use the other two colors, both combinations eg O,G and G,O

**If only one color is present (eg B,B)**
- The three other color pairs (eg G,G or R,R, or O,O)
The only allowed transitions

Reverse Colors

Other two colors (both orientations)

Any other transitions would require the outer two bases to change
These two color changes are not allowed, as 2 bases change
The diagram illustrates the allowed and forbidden transitions for the sequence T.C.A.T.G. Each base is represented by a color:

- **Green** for A
- **Orange** for C
- **Red** for G
- **Yellow** for T

1. **T.C.A.T.G**
   - Allowed 3 colors changed
2. **T.A.G.T.G**
   - Allowed 2 colors changed acceptable "transition"
3. **T.C.G.T.G**
   - Allowed 2 colors changed acceptable "transition"
4. **T.C.G?G**
   - Not allowed 2 colors changed but a forbidden "transition"
Why leave color space?
Why leave color space?
Align color space reads against color space reference
Why leave color space?
Align color space reads against color space reference

SNP 2 colors change
Why leave color space?
Align color space reads against color space reference

Incorrect call, single change in color space
For Research Use Only. Not for use in diagnostic procedures.

Applera, Applied Biosystems and AB (Design) are registered trademarks of Applera Corporation or its subsidiaries in the U.S. and/or certain other countries.