

AMC Hybridization Sample Naming Schema

This document describes how the AMC names the hybridizations run in the lab.

- Explanation of hybridization names
- Single letter codes used to identify chip types
- Special cases

Explanation of Hybridization Name

To accurately track the labeled target and chip subtype used for each hybridization, the AMC assigns each hybridization a unique name. This name serves three purposes: it allows the AMC to identify the sample and its labeling history; it assigns the investigator-designated sample name to each hybridization; and it associates each hybridization with a particular project. Below is a figure that visually shows how a hybridization's name is generated.

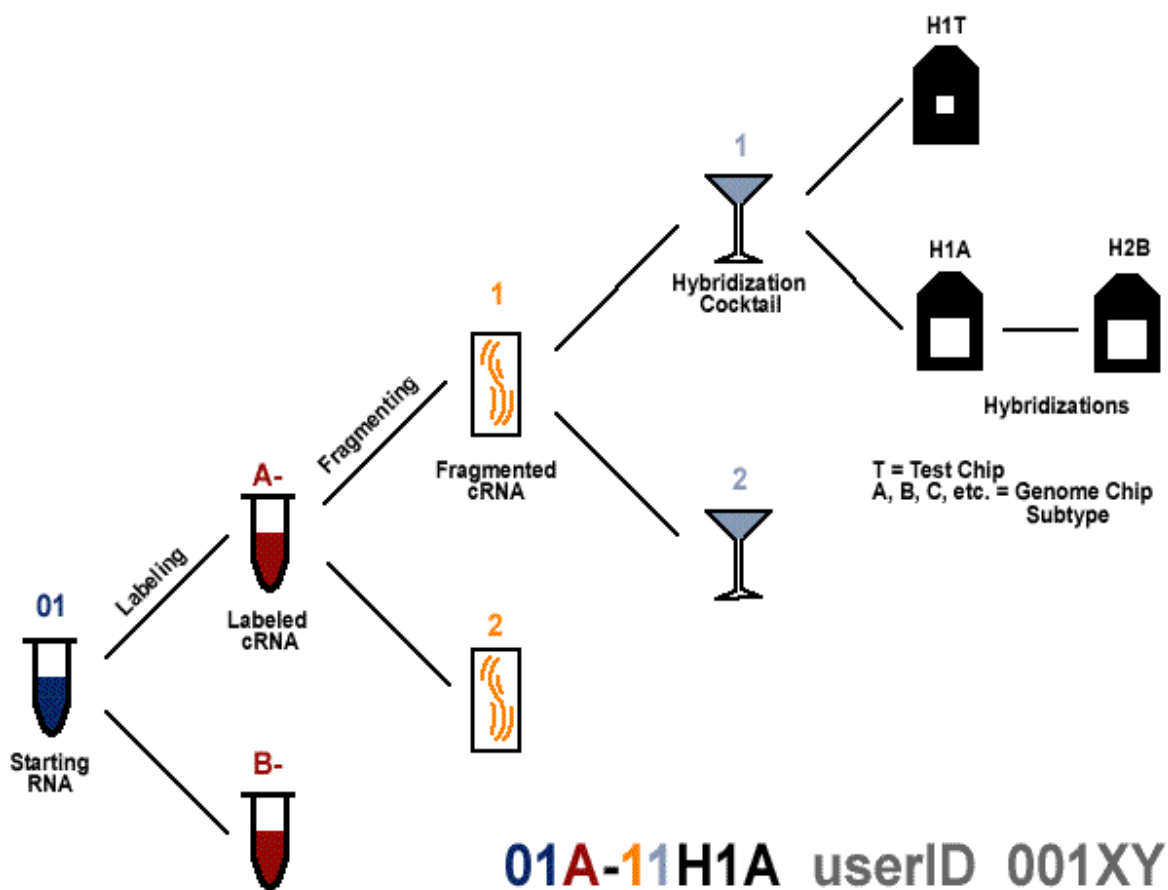


Figure: Visualization of naming schema

Following is an example of a single hybridization name. The name is divided into three parts, each of which is separated by an underscore (_).

01A-11H1A_userID_001XY

Figure: Name example

The part on the far right represents the Project ID for the sample applied to the chip. When accepted by the AMC, each project is assigned a unique ID. Each hybridization run for a project is associated with this project ID. Project ID's takes the form of "001XY" where the "001" is a number assigned by the AMC and the "XY" are the initials of the project's principal investigator.

To give an investigator the ability to easily identify a sample, the AMC provides the opportunity for the investigator to assign each sample a unique 6 character alphanumeric label. This 6-character label is listed in the middle section of the hybridization name. The figure above represents this section as "userID."

The final section of a hybridization's name is dense with information about the sample itself. This part of the name contains information about the sample and its labeling preparation history. The figure below shows the information breakdown for this part of the name.

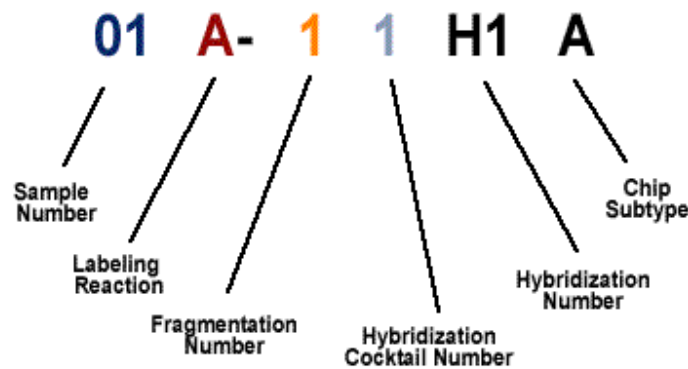


Figure: Name Breakdown

The first two digits identify the original sample used in the labeling protocol. A set of samples in a project are labeled by starting at 01 and continuing, 02, 03, 04, etc. The next character identifies which labeling reaction was used to assemble the hybridization cocktail. The first labeling reaction is designated "A"; if it is necessary to relabel the original sample, the second reaction will be designated "B" and so on. The labeling process produces biotinylated RNA referred to as target cRNA. Prior to hybridization, the cRNA is fragmented. Each fragmentation reaction is assigned a number (1,2 etc.) This number is listed immediately after the hyphen in the name. Multiple hybridization cocktails can be made from a single fragmentation solution so, the hybridization cocktail number is indicated by the numeral to the right of the fragmentation number. A single hybridization cocktail can be hybridized multiple times. The "H*" section of the name tells one how many times this hybridization cocktail has been hybridized. For example, "H2" means this is the second time the cocktail has been hybridized. The last character in this section of the name tells someone what chip subtype is being hybridized (See section on chip subtype codes).

Chip Type Identification

Starting with the addition of a new human chip, some species will have multiple chips that could be designated as "A" or "B" in the naming schema. For example a sample could be run on an HG_U95Av2 chip or a HG_U133A chip and still be named 01A-11H1A_userID_001XY. To avoid any possible confusion arising from this situation, the AMC identifies new chips for a species with a letter other than A,B, etc. For the HG_U133 set, the HG_U133A chip is designated with an "F" in the name schema and the HG_U133B chip is designated with a "G" For purposes of clarity, any time a species becomes represented by a new set of chips, the lettering of the new chips will start with "F" Because of possible confusion between the letter "I" and the number 1, we will not use "I" in the new schema. "T" is reserved for test chips. Below is a listing of the chip codes:

Human Chip Code	
HG_U95Av2	A
HG_U95Bv2	B
HG_U95Cv2	C
HG_U95Dv2	D
HG_U95Ev2	E
HG_U133A	F
HG_U133B	G
HG-Focus	H
"I" Skipped	I
HG-U133A_2	J
HG-U133_Plus_2	K

Murine Chip Code	
MG_U74Av2	A
MG_U74Bv2	B
MG_U74Cv2	C
MOE-430A	F
MOE-430B	G

Rat Chip Code	
RG_U34A	A
RG_U34B	B
RAE-230A	F
RAE-230B	G
Other Chip Codes	
Test 3	T

Special Cases Applied to the Schema

At times, certain procedures are performed on chips that need to be reflected in the sample's name. Since these procedures are done infrequently, they are not a part of the normal schema name and are only added when needed. The procedures are represented by a lower case 2 letter code that follows the AMC ID (01A-11H1A) section of the name. The different procedures and their code are listed below.

Procedure	Code
Rewash a chip with a round of non-stringent washes in the fluidics station after the normal	rw

protocol has been completed	
Restain the chip after it's initial staining protocol	rt
Rescan the chip	rs

For example, If we rewashed a chip because it was showing high background, the name would change from 01A-11H1A_userID_001XY to 01A-11H1Arw_userID_001XY. This allows us to distinguish between the two.

Under normal circumstances, the affymetrix lab database fills a sample's EXP info comment field with just the name of the sample (i.e. 01A-11H1A_userID_001XY). However, in the case of bad data and/or the above special cases, further comment is needed. Therefore, if data is bad and/or the samples have been rewashed, restained or rescanned, the affymetrix staff will enter further information into the EXP info's comment field. Following the sample name, the staff member will enter whether the chip's data is "bad" (not usable) or "good" (usable) followed by any comments about the sample. These notes should be comma delimited. In the case of a wash, restain or rescan, the amc staff will check to be sure the proper sample name is displayed in the comments field.

Ex:

Comments: 01A-11H1A_userID_001XY , bad, high background caused very poor performance

OR

Comments: 01A-11H1Arw_userID_001XY, good, rewashed chip to lower background. This produced desired results.

The separation of sample name, good/bad, comments with commas is important as the AMC database includes these comments when loading sample information and data.

Version 4.2
BT/CAH/KV