RDML discussion thread before the launch of this discussion group

Orginal announcement of RDML-proposal [2005-10-13]

Dear qPCR specialists
Dear friends

At the 2nd International qPCR Symposium in Weihenstephan, Germany, Hellemans et al. proposed a universal XML based data format for exchange and publication of qPCR data.
At http://medgen.ugent.be/rdml/ you can find the details of the proposal, as well as XML schema definitions for the data files. For the non-XML skilled people, there is also a readable RDML layout description the XML file types.

May we kindly ask you to go over the proposed format and provide as much feedback or comments as possible. If you agree with the need and concept, complete the support form on the RDML site (http://medgen.ugent.be/rdml/support.php). By doing this, your name will be listed at the bottom of the RDML home page, indicating you are supporting it.

Please forward this message to other interested colleagues or specialists in the field. In a first phase, we would like to include only academic supporters. Once a critical mass is reached and some agreement is reached on the XML data exchange format, we would send out our proposal to the companies.

Best wishes

Jan Hellemans & Jo Vandesompele

Responses to the proposed RDML format

Looks great. I see you included a lot, e.g. the sample specific efficiency. I will go through the RDML layout and will come back to you with my suggestions. First remarks:
ï¿½ PCR format: ROTOR: there are multiple rotor formats in RG 32 and 72 and new 100 in the RG 6000

> has been added

ï¿½ PCR format: BLOCK: 32, 96, 384

> has been added
ï¿½ Some people still doing 4-step PCR to circumvent detection of primer-dimers, you should include the option.

> has been added

ï¿½ Further you should add how the background fluorescence was calculated, this is due to the used platform.
has been added
ï¿½ CT / CP / TOP (TOP = take off point in RG systems)

> Ct, Cp and TOP have been changed bu a more universal name, i.e. QC or Quantification Cycle (the cycle at which you do the quantification)

--

Good initiative! (We) looked at it and we both find it worthwhile, it will all people talk one language. However, we thought that the application could benefit from (I hope we didn't miss it between all the details):
ï¿½ User name - the person logged on the machine while producing the results.
ï¿½ Time and date the data was created.
ï¿½ Serial number of the machine. This is not the name of the model, but the serial number the machine receives when it leaves the factory (please let me know if I was not clear).

> RDML aims at the minimal required information to process, exhange and re-analyse qPCR data. In this respect, suggestions 1-3 seem not very well suited. Of interest to note is that the language is XML and therefore extensible by nature. So you guys could add additional information internally.

ï¿½ Replicates identification. We think that being able to identify different types of replicates is very useful for later data analysis. Particularly if they contain also information on the sequence of production (i.e. RT is done before PCR).

> RDML defines (PCR) replicates as wells having the same sample name AND the

same gene (assay) name. When using RT or biological replicates,the samples should have a different name, and normalized independently. The purpose of the experiment (incl. any biological replicates) should be indicated in the experiment annotation field.

--

As to the content - I am wondering if a point to address is the quality if the RNA?? Ask for a RIN number or if the quality was assayed? and a relative assessment if the quality?? or some thing of that nature.

> This has been included now.

--

I greatly appreciate the opportunity to contribute. Overall, this initiative has obvious and substantial utility. Inclusion of such extensive data such as Fc (fluorescence) readings, is an exciting development that would greatly improve the currently state of quantitative real-time PCR.

ï¿½ One significant omission is the fluorescence threshold used to calculate Ct values. There is no apparent reason why this should not be included, and indeed an Ft value is required for some types of quantitative calculations. It would be a shame not to have this included, and would be difficult to retrieve otherwise.

>This has been added now. See Run > Run data > Well > Dye > Quantification fluorescence. We have tried to use general names for e.g. Cp/Ct/TOP (Quantification Cycle) and Ft/Threshold (Quantification fluorescence, i.e. the fluorescence intensity at which you determine the quantification cycle to do the calculations).

--

Regarding inclusion of identifying details, I understand the need to simplify the user's life, but I think that at least regarding some of the data there is no contradiction between the two. The user's name (the one who logged on to the machine), time and date should be already in the machine file. This info is part of regulatory requirements and will be useful for commercial data analysis firms.

--

I've been looking at your proposed standard for an XML markup of RT-PCR data. I agree that it would be great to have a standard way of transferring this data from instruments to analysis software. However, I have a few questions/suggestions.

ï¿½ First of all, is the PCR format type going to be extensible? My worry is that, as new instruments come out, they may have different formats than the one included in the standard. You might want to use the idea of a format description, whereby one could describe the format as 96-well where the wells are in a 8x12 plate. This would be able to differentiate this format from, say, the Cepheid SmartCycler instrument, where you can have 16-well blocks that can be chained to be 96 wells, but PCR happens in each well independently of the others. Also, I'm not sure how the SmartCycler's independent PCR conditions would be represented by the rest of the format.

ï¿½ I didn't see anywhere in the standard for a set of standards and their related unknowns. For instance, if you are running multiple experiments on a single plate, there's no way in the format to say that certain standards are associated with certain unknowns and other standards are associated with other unknowns.

ï¿½ Among the sample types, there is one called NAC or "No Amplification Control". What is the definition of an NAC? Is this the same as a positive control? If not, should there be another positive control sample type?

Comments on proposed standard

I've been looking at the proposed standard for an XML markup of RT-PCR data. I agree that it would be great to have a standard way of transferring this data from instruments to analysis software. However, I have a few questions/suggestions.

First of all, is the PCR format type going to be extensible? My worry is that, as new instruments come out, they may have different formats than the one included in the standard. You might want to use the idea of a format description, whereby one could describe the format as 96-well where the wells are in a 8x12 plate. This would be able to

differentiate this format from, say, the Cepheid SmartCycler instrument, where you can have 16-well blocks that can be chained to be 96 wells, but PCR happens in each well independently of the others. Also, I'm not sure how the SmartCycler's independent PCR conditions would be represented by the rest of the format.

I didn't see anywhere in the standard for a set of standards and their related unknowns. For instance, if you are running multiple experiments on a single plate, there's no way in the format to say that certain standards are associated with certain unknowns and other standards are associated with other unknowns.

Among the sample types, there is one called NAC or "No Amplification Control". What is the definition of an NAC? Is this the same as a positive control? If not, should there be another positive control sample type?

Erik

Re: Comments on proposed standard

Dear Erik

Thanks for your valuable feedback. Please find below my responses and get back if something is not clear, or if you have any further suggestions. If possible, please sign with your name at the RDML home page (http://medgen.ugent.be/rdml).

Jo

--- In RDML-format@yahoogroups.com, "erikpaulnielsen" <enielsen@...> wrote:
>
> I've been looking at the proposed standard for an XML markup of RT-PCR
> data. I agree that it would be great to have a standard way of
> transferring this data from instruments to analysis software.
> However, I have a few questions/suggestions.
>
> First of all, is the PCR format type going to be extensible?

By nature it is (X in XML stands for eXtensible). The moment a new format comes along, it will be included (e.g. 1536-well).

My worry
> is that, as new instruments come out, they may have different formats
> than the one included in the standard. You might want to use the idea
> of a format description, whereby one could describe the format as
> 96-well where the wells are in a 8x12 plate. This would be able to
> differentiate this format from, say, the Cepheid SmartCycler
> instrument, where you can have 16-well blocks that can be chained to
> be 96 wells, but PCR happens in each well independently of the others.

This is probably covered, as you have the option to indicate 'single-well' as PCR format if you use the SmartCycler. I think it is not important to know if you chained one or two instruments, the reactions are always individually controlled. Please comment if you don't agree.

> Also, I'm not sure how the SmartCycler's independent PCR conditions
> would be represented by the rest of the format.

Good point. As it is now, the PCR conditions are linked to the entire series of wells in a 'run', not to an individual well. This might be circumvented by exporting a runRDML file for each SmartCycler well. An alternative is to incorporate the PCR conditions (and other info) at the level of each well, but then, runRDML files would probably grow prohibitively large (especially for 384-well formats).

> I didn't see anywhere in the standard for a set of standards and their
> related unknowns. For instance, if you are running multiple
> experiments on a single plate, there's no way in the format to say
> that certain standards are associated with certain unknowns and other
> standards are associated with other unknowns.

Each well has amongst others a type identifier (e.g. unknown, standard, ...), a sample name, and a target (gene/assay) name. The idea would be that wells with the same target name would be considered as a 'set' as you call it (so including both the standards and the unknowns).

> Among the sample types, there is one called NAC or "No Amplification
> Control". What is the definition of an NAC?

This is a negative control. Most often, it is a minus-RT control (using (DNase treated) RNA as template instead of cDNA).

> Is this the same as a

> positive control? If not, should there be another positive control
> sample type?

For now, there is no labeling of a possitive control. This could be added of course. I haven't seen this term in other qPCR instrument software labeling, did you?

> Erik
>

Re: [RDML-format] Re: Comments on proposed standard

```
>  Also, I'm not sure how the SmartCycler's independent PCR conditions
> would be represented by the rest of the format.

Good point. As it is now, the PCR conditions are linked to the entire
series of wells in a 'run', not to an individual well. This might be
circumvented by exporting a runRDML file for each SmartCycler well. An
alternative is to incorporate the PCR conditions (and other info) at
the level of each well, but then, runRDML files would probably grow
prohibitively large (especially for 384-well formats).
```

One technical way to get around this is to have the option of more than one "PCR settings" block with a unique identifier. This identifier could then be included with a sample. This would mean that you only have to list each unique "PCR settings" block once, and the only thing that's repeated is the identifier. If there is only one settings block, it can be assumed that it applies to all samples, and thus the identifier would not be necessary in that case.

```
> I didn't see anywhere in the standard for a set of standards and their
> related unknowns.  For instance, if you are running multiple
> experiments on a single plate, there's no way in the format to say
> that certain standards are associated with certain unknowns and other
> standards are associated with other unknowns.

Each well has amongst others a type identifier (e.g. unknown,
standard, ...), a sample name, and a target (gene/assay) name. The
idea would be that wells with the same target name would be considered
as a 'set' as you call it (so including both the standards and the
unknowns).
```

Unfortunately, this doesn't work if you're, for example, comparing two different primer sets for the same target in the same plate. You would have to name the target differently in this case, unless I'm missing something. Or if you're comparing two different sample preparation methods. The sample and target names would be the same, but the standards for one set (or experiment, if you prefer) should not be used for the other. I still think some other form of

grouping is necessary.

```
> Is this the same as a
> positive control?  If not, should there be another positive control
> sample type?

For now, there is no labeling of a possitive control. This could be
added of course. I haven't seen this term in other qPCR instrument
software labeling, did you?
```

EraGen's software (which is the Plexor software, developed for Promega), has a positive control. It can be used both to indicate an internal positive control that indicates PCR happened in the well or as a positive sample control (also to indicate PCR happened). This is important for clinical assays, as it says in the Draft Guidance on *Nucleic Acid Based In Vitro Diagnostic Devices for Dectection of Microbial Pathogens* that "If the assay system also includes and internal control (recommended by CLIA for test systems incorporating amplification to detect inhibition)...". I believe that probably goes doubly for qPCR assays, at least from a clinical perspective. I don't know if there is a European counterpart to this recommendation. Also, I believe positive internal controls could be used in forensics applications and other assays where you want to make sure nothing went wrong.

Erik

Re: [RDML-format] Re: Comments on proposed standard

Hi all,

In regards to the control types, the MyiQ and iQ5 as well as the MX3005 software platforms allow the user to define positive, negative and no template controls. I know that Bio-Rad included this ability because it was asked for from the researchers, and I can assume that this is why Stratagene also has this ability. I am not as familiar with the other software platforms, so I do not know if Corbett, Roche, Eppendorf and ABI also follow this paradigm.

I think you should have the ability to import several different control types - if you want this to be universal.

Best regards,
Hilary

"Jo Vandesompele" <jvdesomp@...>
Sent by: RDML-format@yahoogroups.com

02/19/2006 01:06 AM

Please respond to

To RDML-format@yahoogroups.com
cc
Subject [RDML-format] Re: Comments on proposed standard

RDML-format@yahoogroups.com

Dear Erik

Thanks for your valuable feedback. Please find below my responses and get back if something is not clear, or if you have any further suggestions. If possible, please sign with your name at the RDML home page (http://medgen.ugent.be/rdml).

Jo

--- In RDML-format@yahoogroups.com, "erikpaulnielsen" <enielsen@...> wrote:
>
> I've been looking at the proposed standard for an XML markup of RT-PCR
> data. I agree that it would be great to have a standard way of
> transferring this data from instruments to analysis software.
> However, I have a few questions/suggestions.
>
> First of all, is the PCR format type going to be extensible?

By nature it is (X in XML stands for eXtensible). The moment a new format comes along, it will be included (e.g. 1536-well).

My worry
> is that, as new instruments come out, they may have different formats
> than the one included in the standard. You might want to use the idea
> of a format description, whereby one could describe the format as
> 96-well where the wells are in a 8x12 plate. This would be able to
> differentiate this format from, say, the Cepheid SmartCycler
> instrument, where you can have 16-well blocks that can be chained to
> be 96 wells, but PCR happens in each well independently of the others.

This is probably covered, as you have the option to indicate 'single-well' as PCR format if you use the SmartCycler. I think it is not important to know if you chained one or two instruments, the reactions are always individually controlled. Please comment if you don't agree.

> Also, I'm not sure how the SmartCycler's independent PCR conditions
> would be represented by the rest of the format.

Good point. As it is now, the PCR conditions are linked to the entire series of wells in a 'run', not to an individual well. This might be circumvented by exporting a runRDML file for each SmartCycler well. An alternative is to incorporate the PCR conditions (and other info) at the level of each well, but then, runRDML files would probably grow prohibitively large (especially for 384-well formats).

> I didn't see anywhere in the standard for a set of standards and their
> related unknowns. For instance, if you are running multiple
> experiments on a single plate, there's no way in the format to say

> that certain standards are associated with certain unknowns and other
> standards are associated with other unknowns.

Each well has amongst others a type identifier (e.g. unknown,
standard, ...), a sample name, and a target (gene/assay) name. The
idea would be that wells with the same target name would be considered
as a 'set' as you call it (so including both the standards and the
unknowns).

> Among the sample types, there is one called NAC or "No Amplification
> Control". What is the definition of an NAC?

This is a negative control. Most often, it is a minus-RT control
(using (DNase treated) RNA as template instead of cDNA).

> Is this the same as a
> positive control? If not, should there be another positive control
> sample type?

For now, there is no labeling of a possitive control. This could be
added of course. I haven't seen this term in other qPCR instrument
software labeling, did you?

> Erik
>