DLBCL Cell of Origin RUO Classifier Design Review (FDR2/3)

DLBCL Team

09-SEP-2014

HTG Molecular

Simple | Accurate | RNA

Agenda

- Overview of the design
  - Classifier algorithms
  - Classifier software
- Verification and validation
  - Classifier functional verification
  - Classifier system validation
  - Software verification and validation
- Approval for release to production

Project goals

- an HTG Edge plate-based assay for research use only (RUO) to measure gene expression in specimens previously typed as DLBCL.

- a classifier algorithm and software to interpret the expression values and classify DLBCL specimens into ABC, GCB or Unclassified group.

Expected products

- qNPA Edge plate-based assay either in a quarter or full plate format with defined genes (including HK and control genes)

- a classifier software, incorporated with the Edge Host System, that performs quality control and classifies DLBCL samples into ABC, GCB or Unclassified.

Two phases

- CP1

  Develop and implement EDGE-based statistical classifier (in R)

  Initial assessment of the classifier's performance with training FFPE samples by cross-validation

  Assess the EDGE2EDGE, run2run variability impact

- **CP2/CP3**

  Re-train and obtain the final classifier algorithms

  Verify the classifiers' performance with independent FFPE samples

  Classifier software development and validation

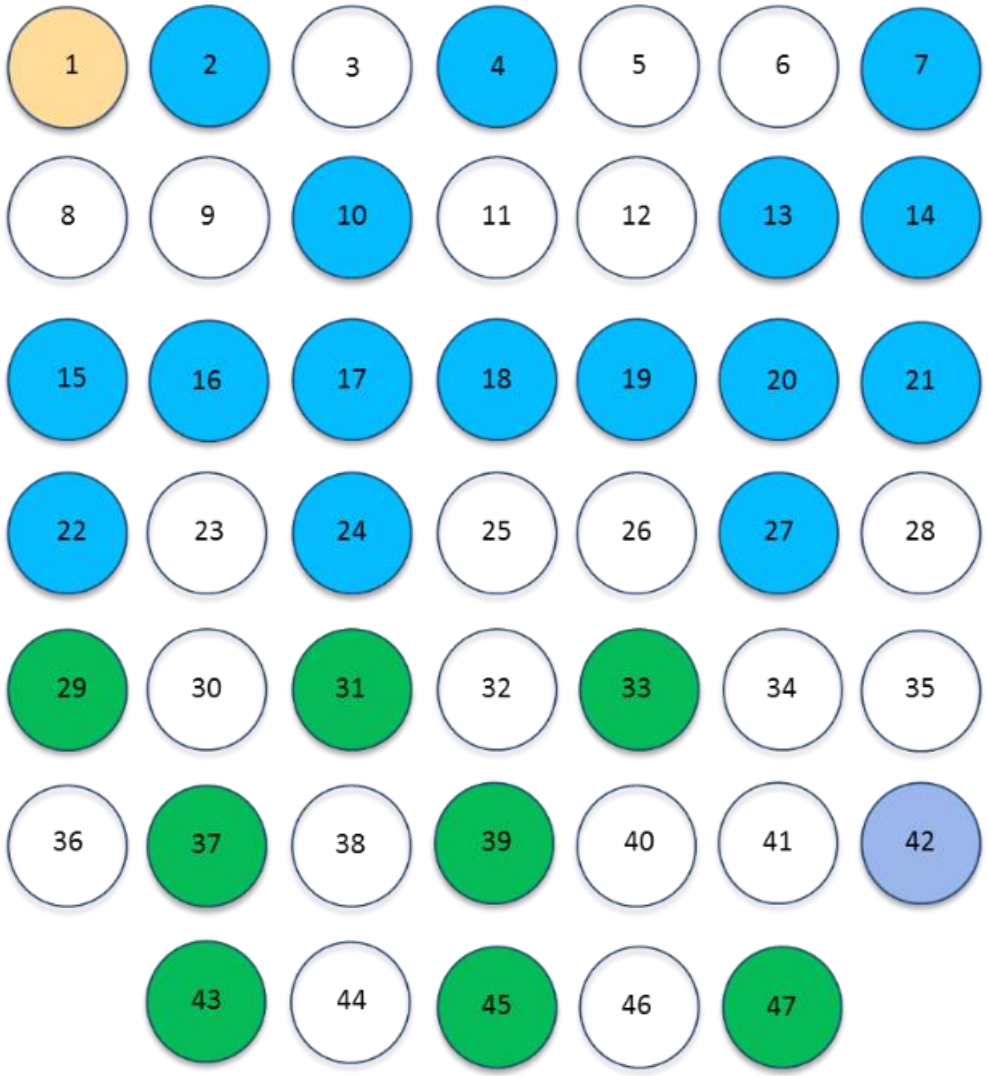
  Integration with EDGE host

  System validation

Overview of the design

Classifier algorithms

HTG DLBCL RUO Array Design (816-093)



26 occupied spots:
16 genes
8 HKs
2 controls

Position	Gene	Position	Gene
1	Pos1	25	
2	BCL6	26	
3		27	MS4A1
4	BCL2	28	
5		29	RPL4
6		30	
7	LMO2	31	RPL19
8		32	
9		33	RPS29
10	FOXP1	34	
11		35	
12		36	
13	MME	37	ACTB
14	LRMP	38	
15	ITPKB	39	TBP
16	MYBL1	40	
17	PIM1	41	
18	IL16	42	ANT
19	CCND2	43	PPIA
20	FUT8	44	
21	IRF4	45	PRKG1
22	ENTPD1	46	
23		47	GAPDH
24	SERPINA9		

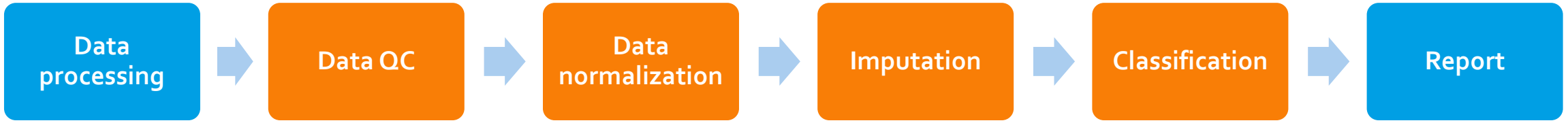
HTG Molecular
Simple | Accurate | RNA

Genes included in the DLBCL COO assay

DLBCL Classification genes		Additional Markers	Housekeeping Controls	Controls
BCL6	FOXP1	CD20 (MS4A1)	RPL4	ANT
LMO2	ENTPD1 (CD39)	BCL2	RPL19	POS1
MME (CD10)	CCND2	IL16	RPS29	
LRMP	FUT8	PIM1	GAPDH	
ITPKB	IRF4		ACTB	
MYBL1			TBP	
SERPINA9 (GCET)			PRKG1	
			PPIA	
12 genes		4 genes	8 genes	2 genes

HTG Molecular
Simple | Accurate | RNA

Overview of DLBCL COO classifier workflow



- .csv files generated from EDGE software

- Process QC
- Well QC
- Sample QC

- HK normalization

- Impute missing classifier genes

- Classify samples into ABC, GCB or Unclassified

- Output sample/well QC status and classification result

Metrics or algorithms have been developed for each components

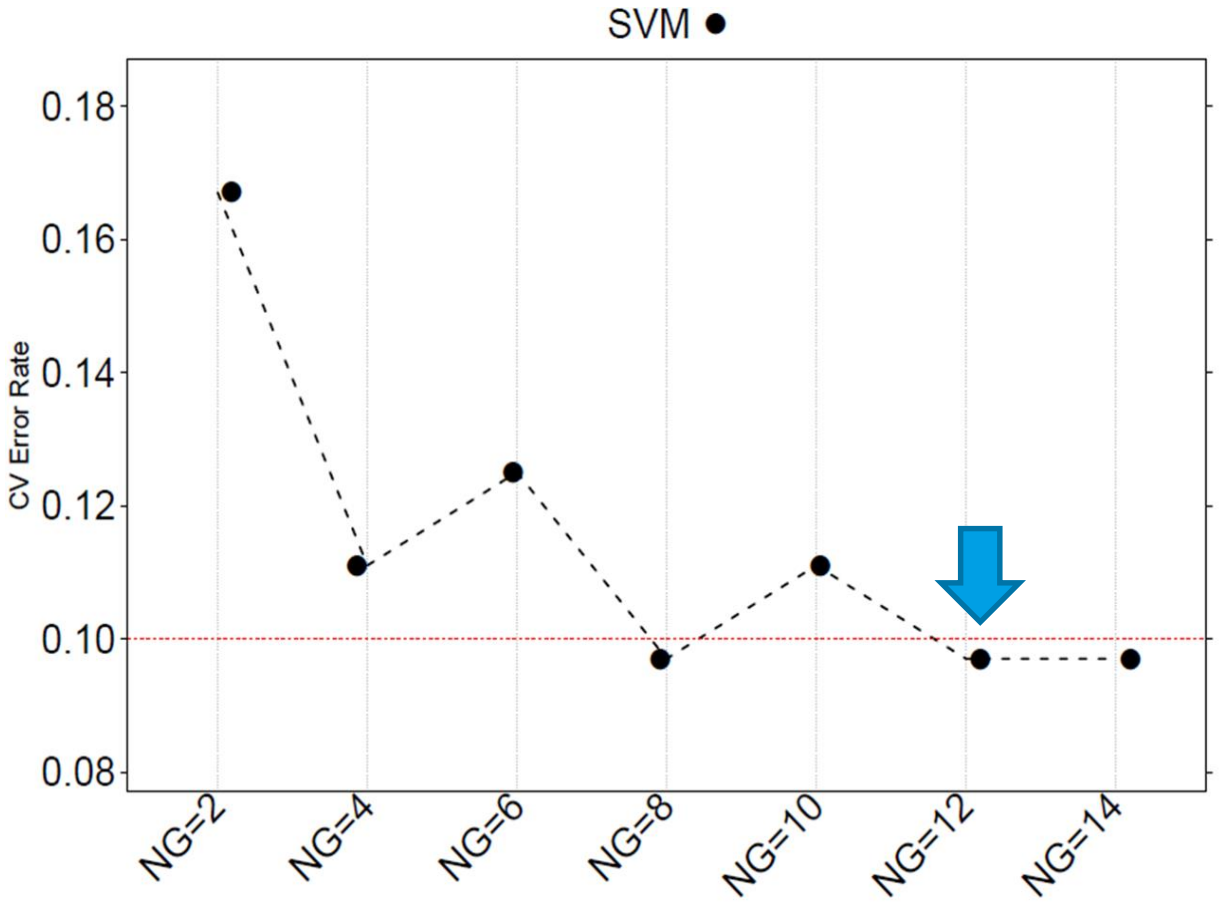
QCs and imputation

- Outlier gene removal
  - Only remove one of the three replicated genes (if there are 3 replicates)
  - Does not remove well.
- Well/replicate removal
  - IVT/ANT (2000, 1200)
  - HK lower bound to remove well due to QNS
  - HK upper bound to remove well due to Saturation
  - Replicate variability QC by regression
- Sample removal
  - If <2 replicates left after QC, samples are removed.
- After QC, samples are subject to "missing value imputation"
  - Can impute missing value if 1 classifier gene missing

Classification algorithms used during classifier training

- Explored multiple statistical algorithms
  - SVM
  - Bayesian model
  - RF
  - LDA/QDA
  - Logistic regression
- Cross-validation for method performance evaluation and error estimate

Gene numbers and CV error rate



- 16 candidate genes on the array
  - BCL2 and MS4A1 not considered due to they are not associated with classification biologically
- 14 genes used for classifier training
  - 2 additional genes(IL16 and PIM1) were removed due to lack of class separation power
- 5 HKs
  - 3 were removed due to low/missing intensity and high variability

Genes included in the DLBCL COO Classifier

DLBCL Classification genes			Housekeeping Controls	Controls	
BCL6	FOXP1		RPL4	ANT	
LMO2	ENTPD1 (CD39)		RPL19	POS1	
MME (CD10)	CCND2		RPS29		
LRMP	FUT8		GAPDH		
ITPKB	IRF4		ACTB		
MYBL1					
SERPINA9 (GCET)					

12 genes

5 genes

Classifier training and re-training

- Motivation
  Initial classifier with error rate ~19% on the first set of Verification data
  Decision to re-train the classifier by rolling in the Verification data

- 114 candidate sample data available for retraining
  69 from original training
  45 from Verification
  7 MDA samples were removed due to potential label issues
  -> 107 samples used for building model

	ABC	GCB	Total
CHTN	11	5	16
JWCI	5	5	10
MDA	42	39	81
Total	58	49	107

- Two statistical algorithms were top performers
  SVM
  Bayes rule

Classification training CV error rate and cutoff value

Cutoff band	TotalSample	SVM.error.rate	UC by HTG-SVM	# of samples used-SVM	Bayes.error.rate	UC by HTG-Bayes	# of samples used-Bayes
0.1-0.9	107	0.0149	40	67	0.0217	15	92
0.15-0.85	107	0.0506	28	79	0.0313	11	96
0.2-0.8	107	0.0556	17	90	0.0404	8	99
0.25-0.75	107	0.0745	13	94	0.0400	7	100
0.3-0.7	107	0.0990	6	101	0.0583	4	103
0.35-0.65	107	0.0971	4	103	0.0583	4	103
0.4-0.6	107	0.0952	2	105	0.0762	2	105

Cutoff values derived based on cross-validated error rate and
Trade-offs between error rate and call rate

Summary of classifier training

- Developed comprehensive QC metrics
- Develop missing value imputation
- Derived two 12-gene classifiers
  - SVM
  - Bayes

Overview of the design

Classifier software

Two versions of classifier software by design

- Functionality
  - Perform quality control
  - Classify DLBCL samples into ABC,GCB or Unclassified
  - Current version of the classifier is SVM-based

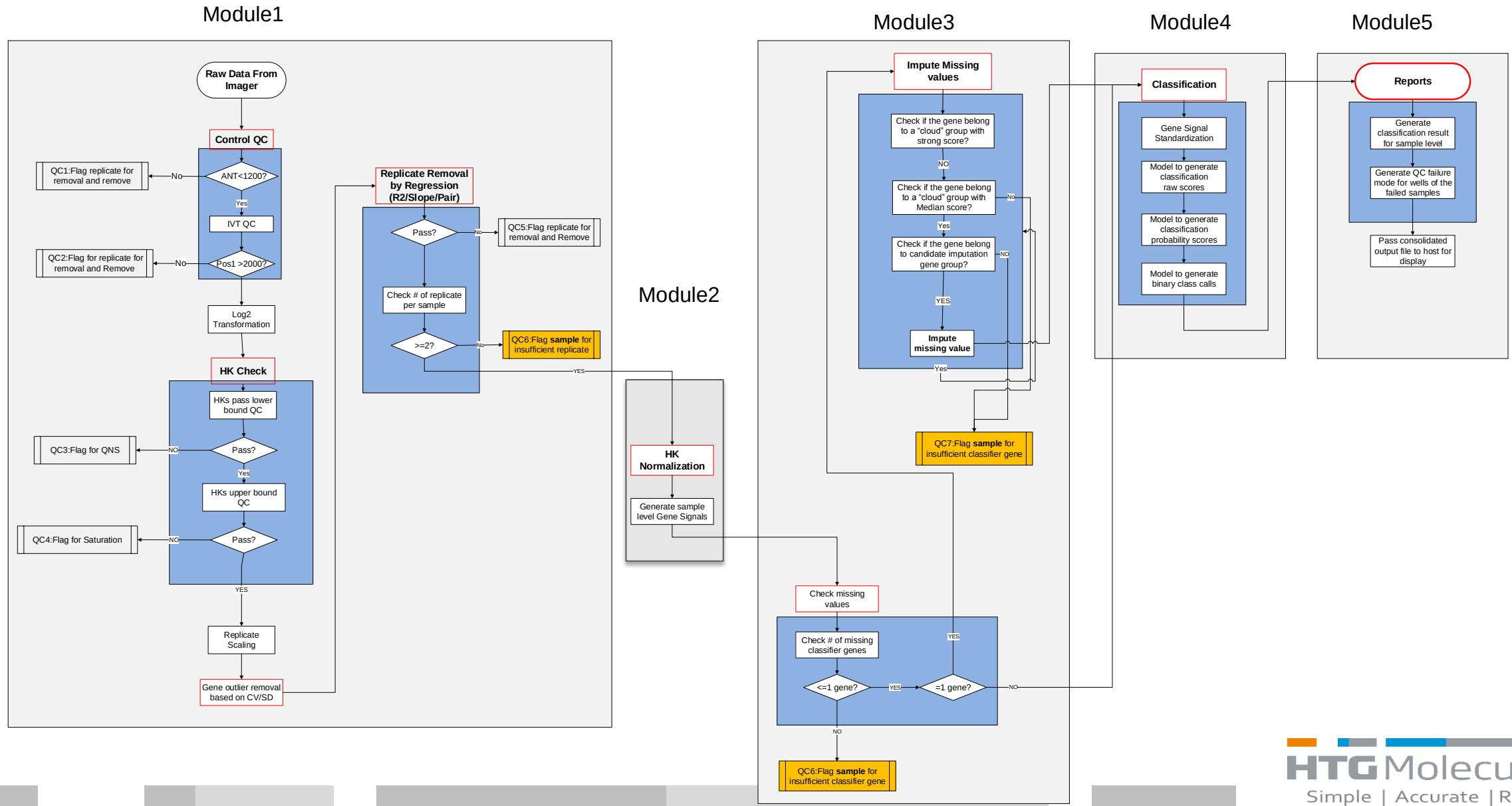
- Command line standalone software
  - Linux
  - Windows (.exe)

- .exe incorporated with the Edge Host System
  - Windows

Classifier software design

- The software consists of five modules
  - QC
  - HK normalization
  - Imputation
  - Classification
  - Report of results
- Seven(7) QC types
  - QC1-7
- Three (3) classification classes
  - ABC, GCB or Unclassified

Classifier software design



An example of classifier software (.exe) output

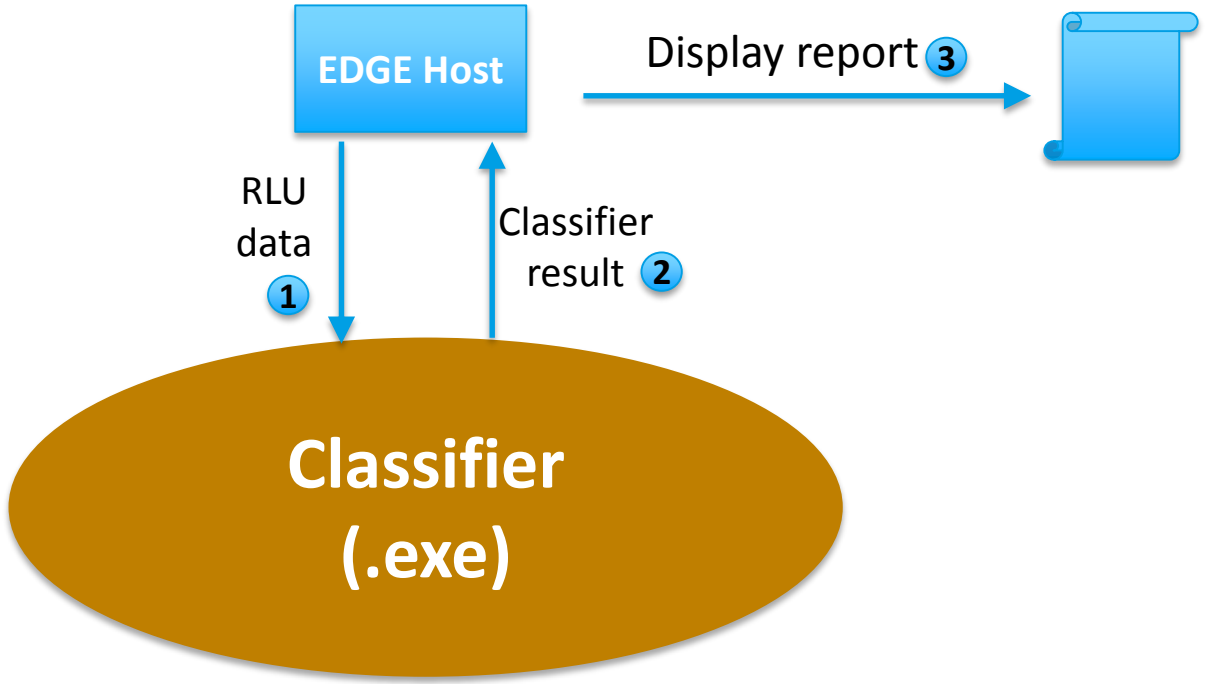
Classifier software .exe output a .csv file containing the following information

SampleID	Results
S1	DNP;A:2=QC1;A:1=QC3;A:3=QC3;
S6	DNP;F:3=QC1;F:1=QC3;F:2=QC3;
S7	DNP;G:2=QC1;G:1=QC3;QC6;
S8	DNP;H:1=QC1;H:2=QC3;H:3=QC3;
S9	ABC
S2	GCB
S5	GCB
S10	GCB
S16	GCB
S3	Unclassified
S13	Unclassified
S15	Unclassified
S17	Unclassified

Code	Details
DNP	Did Not Pass
QC1	Excessive signal from the negative control
QC2	Insufficient signal from the positive control
QC3	Insufficient signal obtained from the HK genes
QC4	Excessive signal obtained from HK genes
QC5	Excessive variability in the replicates
QC6	Insufficient wells for a sample (1 rep)
QC7	Insufficient classifier gene count

Integration with Edge host

1. Edge host passes the RLU data stored in a .csv file to classifier software (.exe)
2. The classifier software (.exe) reads the .csv file and performs
   - QC -> normalization/scaling -> imputation-> classification
   - -> output the result as a .csv file
3. Edge host reads in and displays the classifier result



Edge host display

DLBCL Classifier Results

Research Use Only. Not for use in diagnostic procedures.

HTG Edge System

Printed: 9/3/2014 3:31 PM

Sample Plate Details

Sample Plate Barcode:	23026200152105	**Replicate Format:**	Triplicate
Plate Name:	DLBCL Valid Plate2 P35R3 02Sep2014	**Sample Layout:**	Rows
Test Type:	Edge mRNA	**Control Types:**	None
Plate Format:	4/4	**# of Protocols Used:**	1

Processor Details

Run #:	16
Run Start:	9/2/2014 9:11 AM
Name / Serial #:	35
Run Operator:	Admin User
Run Complete:	9/3/2014 12:17 PM

Reader Details

Run #:	3127
Run Start:	9/3/2014 12:24 PM
Name / Serial #:	20003
Run Operator:	Admin User
Run Complete:	9/3/2014 1:05 PM

Classifier Results

Wells	Sample Name	Result
A1,A2,A3	L-9 1:4	ABC
C1,C2,C3	R-7 1:4	ABC
D1,D2,D3	L-26 1:8	ABC
E1,E2,E3	L-52 1:16	GCB
F1,F2,F3	A-25 1:8	ABC
G1,G2,G3	A-43 1:8	ABC
H1,H2,H3	A-7 1:8	GCB
A4,A5,A6	B-16 1:4	ABC
B4,B5,B6	B-37 1:4	ABC

DLBCL Classifier Results

Research Use Only. Not for use in diagnostic procedures.

HTG Edge System

Printed: 9/3/2014 3:31 PM

Classifier Results

Wells	Sample Name	Result
B7,B8,B9	D-22 1:8	ABC
C7,C8,C9	E-08 1:8	ABC
F7,F8,F9	K-24 1:8	GCB
G7,G8,G9	K-25 1:16	ABC
A10,A11,A12	K-34 1:6	ABC
C10,C11,C12	L-12 1:8	GCB
D10,D11,D12	WD20398:0.2	GCB
E10,E11,E12	WD20405:0.2	ABC
F10,F11,F12	Q-12 1:6	ABC
G10,G11,G12	SU-DHL-2:25000	Unclassified
H10,H11,H12	SU-DHL-6:25000	GCB
E4,E5,E6	C-34 1:16	DNP;QC7;
H7,H8,H9	K-3 1:6	DNP;QC7;
B1,B2,B3	Lysis Buffer DLBCL	DNP;B:1=QC3;B:2=QC3;B:3=QC3;
D4,D5,D6	Lysis Buffer DLBCL	DNP;D:4=QC3;D:5=QC3;D:6=QC3;
D7,D8,D9	Lysis Buffer DLBCL	DNP;D:7=QC3;D:8=QC3;D:9=QC3;
E7,E8,E9	WD20402:0.2	DNP;E:7=QC3;E:8=QC3;QC6;
B10,B11,B12	Lysis Buffer DLBCL	DNP;B:10=QC3;B:11=QC3;B:12=QC3;

Classifier verification

Verification cohort

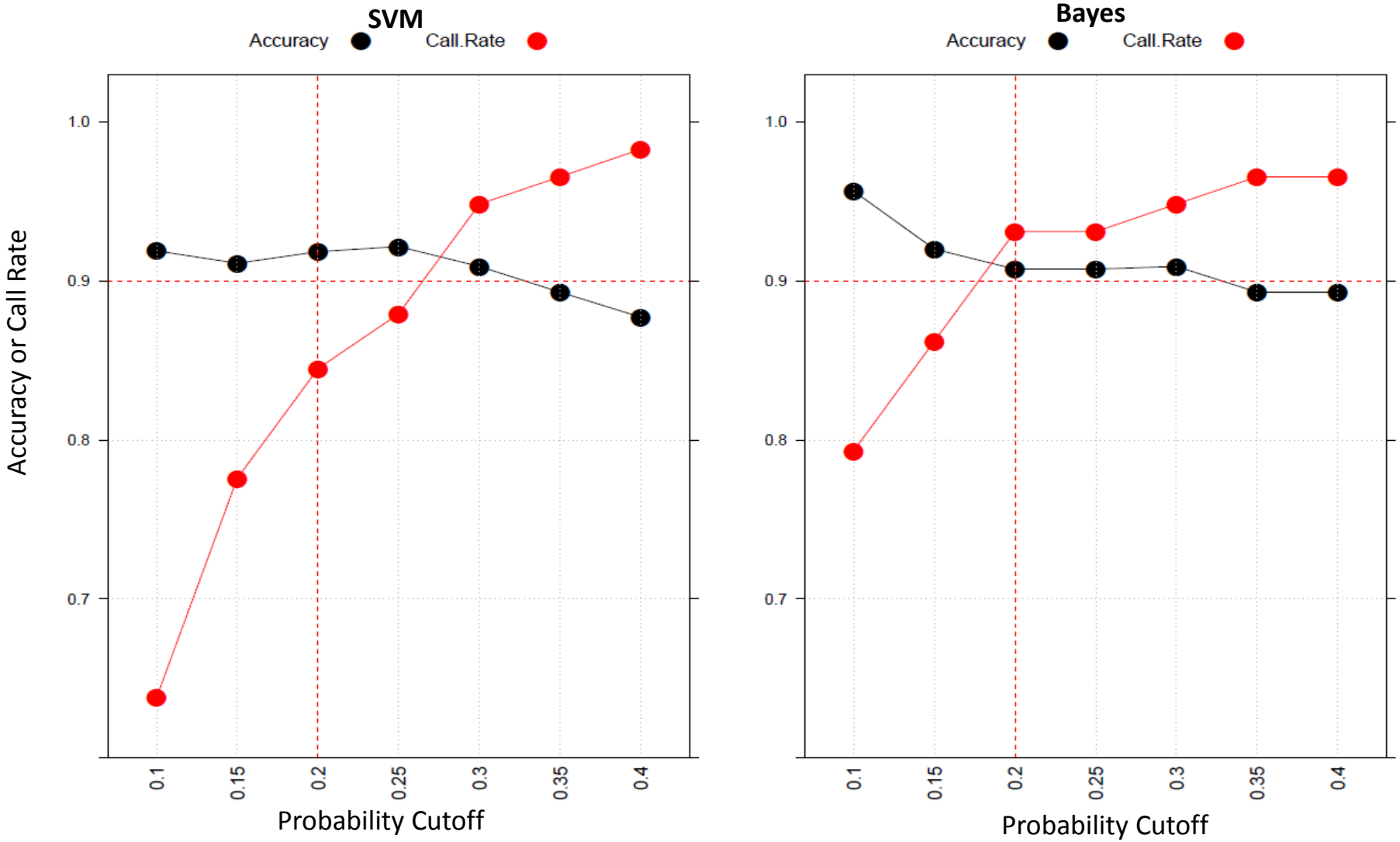
- MDA
  - Total 70 processed
    - *65 pass QC*
    - *7 are MDA GEP UC*
    - *->58 with ABC/GCB MDA GEP calls.*

    - *58 samples were used to calculate concordance*

Verification result summary with MDA cohort

Cutoff band	TotalSample	SVM.error.rate	UC by HTG-SVM	# of samples used-SVM	Bayes.error.rate	UC by HTG-Bayes	# of samples used-Bayes
0.1-0.9	58	0.08	21	37	0.04	12	46
0.15-0.85	58	0.09	14	44	0.08	8	50
0.2-0.8	58	0.08	8	50	0.09	4	54
0.25-0.75	58	0.08	7	51	0.09	4	54
0.3-0.7	58	0.11	2	56	0.09	3	55
0.35-0.65	58	0.11	2	56	0.11	2	56
0.4-0.6	58	0.12	1	57	0.11	2	56

Verification using MDA samples



Classification result met CMRD requirement of >=90% concordance with ref labels

Summary

- We have successfully verified two classifier algorithms
  - SVM and Bayes
- Verification concordance with MDA GEP met CMRD requirement
  - >90% concordance with reference labels

System validation

System validation

- Objective
  - Validate the system performance of the DLBCL Cell of Origin (COO) classifier on Edge system using MDA samples in verification runs that passed QC
- Samples
  - For validation study 57 MDA with GEP labels (ABC/GCB:30/27)
- Acceptance criteria
  - QC criteria
    - If 5 out of the 57 MD Anderson samples failed QC, the EDGE run/validation is considered as a failure.
    - 3 out of the 5 control samples need to be classified correctly and minimum 2 of 3 ABC CHTN controls samples need to be classified correctly for each plate run
  - Concordance criteria
    - Percent concordance between HTG results and Reference labels must be >= 90%
    - Percent concordance between verification and validation testing must be >=90%

Summary of system validation result

- QC and control samples passed acceptance criteria
  - *3 MDA samples failed QC*
  - *all CHTN Samples were classified correctly*
- Concordance with GEP passed acceptance criterion of 90%
  - *Concordance rate is 93.5%*
- Concordance between verification and validation passed acceptance criterion of 90%
  - *Concordance rate is 92.6%*

System validation conclusion

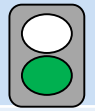
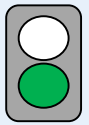
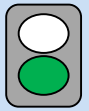
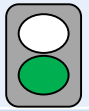
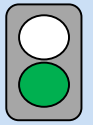
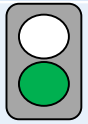
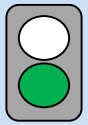
All the acceptance criteria defined in the validation protocol were met !

Software verification and validation

Software validation summary

- Summary report
  - Final release software:
    - *dlbcl.v.1.0.0.00025.exe*
  - Time to complete testing
    - *33 hours*
  - Exception conditions
    - *None*
  - List of issues identified
    - *None*

Software validation testing summary

Test	Description	Software Requirements Specification (SRS)	Pass / Fail
QC	Verify the data have been quality controlled	4.3.1	
Normalization	Verify the data have been HK normalized and technical mean calculated	4.3.2	
Imputation	Verify the missing values have been imputed	4.3.3	
classification	Verify the DLBCL samples have been classifier	4.3.4	
Input and Output Verification	Verify classifier generate the correct output for the proper input	4.3.5	
Options Control	Verify the control options of classifier function properly	4.3.6	
Robustness	Verify the results generated from the classifier are consistent with those from the biostats internal pipeline	4.3.7	

PDP Deliverables

CP2 / CP3 PDP Deliverables

- ☑ Business Plan updated (same as CP1)
- ☑ Commercialization Plan updated (same as CP1)
- ☑ Customer and Market Requirements Document (CMRD) updated (same as CP1)
- ☑ Design and Development Plan (DDP) update
- ☑ Assay FDR/Trace Matrix updated
- ☑ Software Trace Matrix updated
- ☑ Software Development and Validation Plan (SDVP) updated
- ☑ Software Design Document complete
- ☑ Regulatory Strategy and Plan (RP) updated
- ☑ Risk Management Plan updated
- ☑ Hazard Analysis Matrix (HAM) updated
- ☑ dFMEA complete
- ☑ Design Verification and Validation complete
- ☑ Design Transfer complete
- ☑ Service and Support materials complete
- ☑ Design History File (DHF) updated

Next steps

- CP2/3 gate IC approval

Acknowledgement

- Qian Liu
- Cassie Chen
- Grant Schissler (intern)
- Brooke Rabe (intern)
- James Lu

- Kevin Knapp
- Josh Miller
- Jenna Flores-blachtel

- Pat Roche
- David Yu
- Katie Lux
- Mark Schwartz
- Jennifer Kessler
- Peggy Springer

- Sam Rua
- Joely Kremer

- Byron Lawson

- John Luecke

- Don Ball
- Karl Hecker
- Tyler Breedlove

- John Wineman
- Debrah Thompson

Questions??

Thank You!!