Supplementary Methods/Results:

A schematic of the overall experimental design is shown in Figure S5

RNA extraction

20µm was sectioned off the face of the FFPET blocks and discarded. Scrolls of 10µm thickness were cut to a surface area of 1cm^2 – eg if the biopsy surface was 10mm by 5mm, then 2 scrolls were cut. Scrolls were shipped and stored at 4°C until extraction, which occurred within 14 days of scrolling. Nucleic acid extractions were performed independently at the Molecular Characterization Laboratory (MoCha) (FNLCR, Frederick, MD) and the Centre for Lymphoid Cancer (CLC), BC Cancer Agency (Vancouver, Canada). Nucleic acids were extracted using the QIAGEN AllPrep DNA/RNA FFPE Kit (catalogue number 80234, QIAGEN GmbH, Germany), with deparaffinization achieved using the QIAGEN Deparaffinization Solution (catalogue number 19093), according to the manufacturer's instructions. The RNA was quantitated using spectrophotometry (NanoDrop, Thermo Science, DE).

Gene expression

200ng of RNA was used to determine gene expression levels by means of NanoString technology (NanoString Technologies, WA). The total RNA was hybridized to the custom codesets at 65°C overnight (15.5 to 22.5 hours). The reaction was processed on the nCounter™ Prep Station and gene expression data was then acquired on the nCounter™ Digital Analyzer at the "high resolution" setting. At the MoCha (FNLCR, Frederick, MD), "generation 1" nCounter™ machines were used and this setting equated to 600 fields of view,

while at the CLC, “generation 2” machines were used and this setting equated to 280 fields of view.

On the 12 lane NanoString cartridges, 11 RNA samples were run and one synthetic oligonucleotide reference sample. This reference sample consisted of 100nt oligonucleotides representing the targets within the genes that the codeset hybridizes with. These Ultramer® DNA oligonucleotides were manufactured by Integrated DNA Technologies (IA), mixed together and then aliquoted to ensure that the oligonucleotides were in fixed ratios to each other. 5µl of the oligonucleotide reference standard, at a concentration of 1.8pM per oligonucleotide, underwent the same hybridization with the NanoString codeset as the RNA samples – final concentration per oligonucleotide in the hybridization reaction was 300fM.

Standard QC (as offered by nSolver™ Analysis Software, NanoString Technologies, WA) was employed, with flagging of any sample with a total of the positive spike-in controls being outside of 0.3 to 3 times the geometric mean of the total positive spike-in for that cartridge. No samples in these experiments were flagged.

Model Building

The samples used in the training and validation sets were drawn from the previously analyzed set of Lenz *et al*[1] samples for which COO subtype had been estimated according to a Bayesian compound covariate predictor, trained on samples from patients treated with CHOP and validated on a subset of samples from patients treated with RCHOP. At the heart of this prediction was a predictor

score made from a weighted average of the log transformed expression values for genes differentially expressed between ABC and GCB. Given the relatively limited size of the training set, it was decided that rather than form an FFPET predictor *de novo*, the strategy would be to replicate the frozen tissue (FT) "gold standard" predictor scores that had been previously validated in Lenz *et al* using Nanostring technology on FFPET. The same Bayesian derived cutoffs would then be used to divide samples into ABC, GCB and Unclassified categories.

To this end we identified a set of 15 predictive genes for which FT and FFPET measures were found to be well correlated (based on a pilot study of samples independent of the validation set), and which also exhibited a high degree of differential expression between the ABC and GCB subtypes on the CHOP subset of the Lenz *et al* data. Additionally we identified 5 "housekeeping" genes that had low variability across samples in both FT and FPPET (again from the pilot study). The 20 genes are shown in Fig 1A.

The data for each array were normalized by dividing the observed expression values by the geometric mean of the housekeeping genes. These genes were also used to identify poor quality samples. If the geometric mean of the housekeeping genes for a given array was found to be less than 20, the sample was deemed poor quality and was excluded from analysis. To adjust for potential batch effects between runs, a reference array of synthetic oligonucleotides of fixed concentration ratios (see above) was included in each run of patient samples. Loading of the reference array was normalized by dividing the observed expression values by the geometric mean of the

housekeeping genes in that array. The patient samples were then further normalized by dividing the signal values of the patient samples by the corresponding signal values of the reference. After normalization the data was log_2-transformed.

A weighted average of the 15 predictive genes was used to generate a preliminary predictive score. The weight for each gene was based on a combination of the degree of differential expression within the CHOP subset of the Lenz *et al* data, and the correlation between FT and FFPET expression for that gene on the training samples. The preliminary predictor scores on the training samples were found to be highly correlated with their corresponding "gold standard" scores, although with very different scale and centering. To correct this, the scores were linearly transformed so that the mean and variance of the training set FFPET predictor scores equaled that of their corresponding "gold standard" scores. Samples were then divided into ABC, GCB and Unclassified according to the same score cutoffs that were used in Lenz *et al*. Results in the training cohort from both the MoCha and the CLC are shown in the Figures S6 and S7 and Table S3. At MoCha, 46/51 FFPET biopsies in the training cohort yielded GEP that passed the QC criteria, while 49/51 passed at the CLC. The age of the FFPET blocks ranged from 6 to 32 years and the all cases that failed QC were at least 19 years old.

This model, including normalization strategy, criteria for excluding poor samples ("QC"), coefficients, strategy for scaling and centering and cutpoints, was "locked" prior to receipt of the data from the independent validation cohort.

Furthermore, these data were analyzed “blinded” to the “gold standard” COO. The validation set was normalized in the same manner as the training set. A preliminary predictor score was generated using the weights identified on the training set, and then linearly adjusted by the same coefficients that were used to match the mean and variance of the training set FFPET model to that of the training set gold standard. Samples were divided into ABC, GCB and Unclassified samples according to the Bayesian cutpoints identified from the Lenz *et al* training set (Fig S6). Since all of the validation set samples were from the R-CHOP validation group in Lenz, they did not influence the gene selection, model coefficients, or Bayesian cut-points for either the “gold standard” model, or the FFPET model derived on the training set.

Details of the algorithm and methods will be made freely available to academic investigators for research use upon request. Contact the Technology Transfer Center, NCI (9609 Medical Center Dr., Room 1E-530, MSC 9702, Bethesda, MD 20892-9702, Phone: 240-276-5530, Fax: 240-276-5504) to arrange an MTA for research use. A patent governing this method has been filed and commercial entities should contact the Technology Transfer Center, NCI for information regarding a license.

Results and discussion of the “Unclassified” cases by the “gold standard” method

The “unclassified” group has been defined as those tumors where the confidence that they fall into the ABC or GCB subtypes is less than 90%[2] by the “gold standard” method. This assignment has the useful characteristic that it adds robustness to the ABC and GCB designation by reducing the potential for

migration between these 2 groups on repeat testing, migration between platforms or minor changes to the COO model (gene coefficients and thresholds). Thus, in an assay that accurately replicates the original model, it would expected that the frank misassignment from ABC to GCB (or vice versa) would be very low – this is consistent with the 2% (1 out of 57) misassignment seen with the Lymph2Cx assay. The expectation would be that the majority of difference in assignment between the Lymph2Cx and the “gold standard” method would be observed in those considered “unclassified”. When specifically considering this group of 10 tumors by the “gold standard” method, only 2 remain “unclassified” by the Lymph2Cx assay at the MoCha and CLC sites, with the majority being assigned as GCB (see Table S2 and Figures 1 and S2), representing movement around this “middle ground”. Comparisons between the “gold standard” method and the Lymph2Cx scores (Figure S6) reveal a tendency for the scores in the Lymph2Cx assay to be lower than that of the “gold standard” method in the validation cohort – a trend not seen in the training cohort. This tendency is consistent between the 2 Lymph2Cx sites and it is unclear whether this represents bias of the Lymph2Cx assay or of the “gold standard” in the validation cohort. Thus, it is possible that the results in the validation cohort are an under-estimation of the true accuracy of the assay.

Migration in and out of the “unclassified” group between the Lymph2Cx and “gold standard” method occurred in 11 tumors at the MoCha site and 12 at the CLC site. If this migration is included, the absolute concordance between the Lymph2Cx assignments and the “gold standard” is 54 of 67 (82%) at the MoCha

site and 53 of 67 (81%) at the CLC site. The consequences of this migration will ultimately be dependent on how these “unclassified” cases are treated - allowing the physician or clinical trial group to weigh up the risks of an intervention against the uncertainty of the DLCBL subtype in these cases. For example, it may be decided to include these patients in a trial where the intervention has activity selective to a given subtype but minimal side effects but exclude such patients where the intervention has more significant risks. We contend that this is superior to the binary IHC algorithms that hide the uncertainty that exists in these tumors.

Tissue Microarray and Immunohistochemistry

A tissue microarray (TMA) of duplicate 0.6mm cores was constructed from the 60 cases in the validation cohort that had sufficient FFPET available. The TMA was stained, on a Benchmark XT (Ventana, AZ) using the antibodies shown in table S4 and the UltraView Multimer Detection Kit (Ventana, AZ). Scoring of staining in the tumor cells at 10% increments was performed by 2 independent hematopathologists (AM and LMR). Where there was discordance, consensus was reached with a third hematopathologist (RDG). The immunohistochemistry-based COO algorithms were applied as previously described[3-5]. 3 of the 60 cases failed to provide interpretable results.

References:

1. Lenz G, Wright G, Dave SS, et al. Stromal gene signatures in large-B-cell lymphomas. *N Engl J* Med. 2008;359(22):2313-23.
2. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci U S A*. 2003;100(17):9991-6.

3. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 2004;103(1):275-82.
4. Choi WW, Weisenburger DD, Greiner TC, et al. A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. *Clin Cancer Res*. 2009;15(17):5494-502.
5. Meyer PN, Fu K, Greiner TC, et al. Immunohistochemical methods for predicting cell of origin and survival in patients with diffuse large B-cell lymphoma treated with rituximab. *J Clin Oncol*. 2010;29(2):200-7.
6. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N Engl J Med*. 1993;329(14):987-94.

Characteristic		**Training cohort (n = 51)**		**Validation cohort (n = 68)**
Age – median (range)		65 (18 – 83)		61 (16 – 86)
Gender - n (%)				
Male		29 (57)		42 (62)
Female		22 (43)		26 (38)
Stage – n (%)				
I/II		23 (45)		30 (46)
III/IV		28 (55)		35 (54)
missing				3
Lactate dehydrogenase – n (%)				
≤ ULN		20 (47)		24 (44)
> ULN		23 (53)		30 (56)
missing		8		14
ECOG PS – n (%)				
0 – 1		33 (70)		41 (67)
≥ 2		14 (30)		20 (33)
missing		4		7
Extranodal sites – n (%)				
0 – 1		42 (89)		50 (86)
>1		5 (11)		8 (14)
missing		4		10
IPI – n (%)				
0 – 2		28 (61)		43 (69)
3 – 5		18 (39)		19 (31)
not calculable		5		6

Table S1: Patient characteristics of the cohorts.

Percentages given are for patients where the characteristic was available.

Abbreviations: ULN: upper level of normal; ECOG PS: Eastern Cooperative Oncology Group Performance Status; IPI: International Prognostic Index[6]

Antibody	Clone	Manufacturer	Catalogue #	Dilution
BCL6	LN22	Novocastra, Leica Biosystems	NCL-L-BCL-6-564	1:100
CD10	7G3	Novocastra, Leica Biosystems	NCL CD10 270	1:25
FOXP1	JC12	Pierce Endogen	MA1-84005	1:1000
GCET1	RAM341	Abcam	Ab68889	1:50
LMO2	1A9-1	Santa Cruz	sc-65736	1:1000
IRF4/MUM1	MUM1p	Dako	M7259	1:50

Table S4: Antibodies used for immunohistochemistry

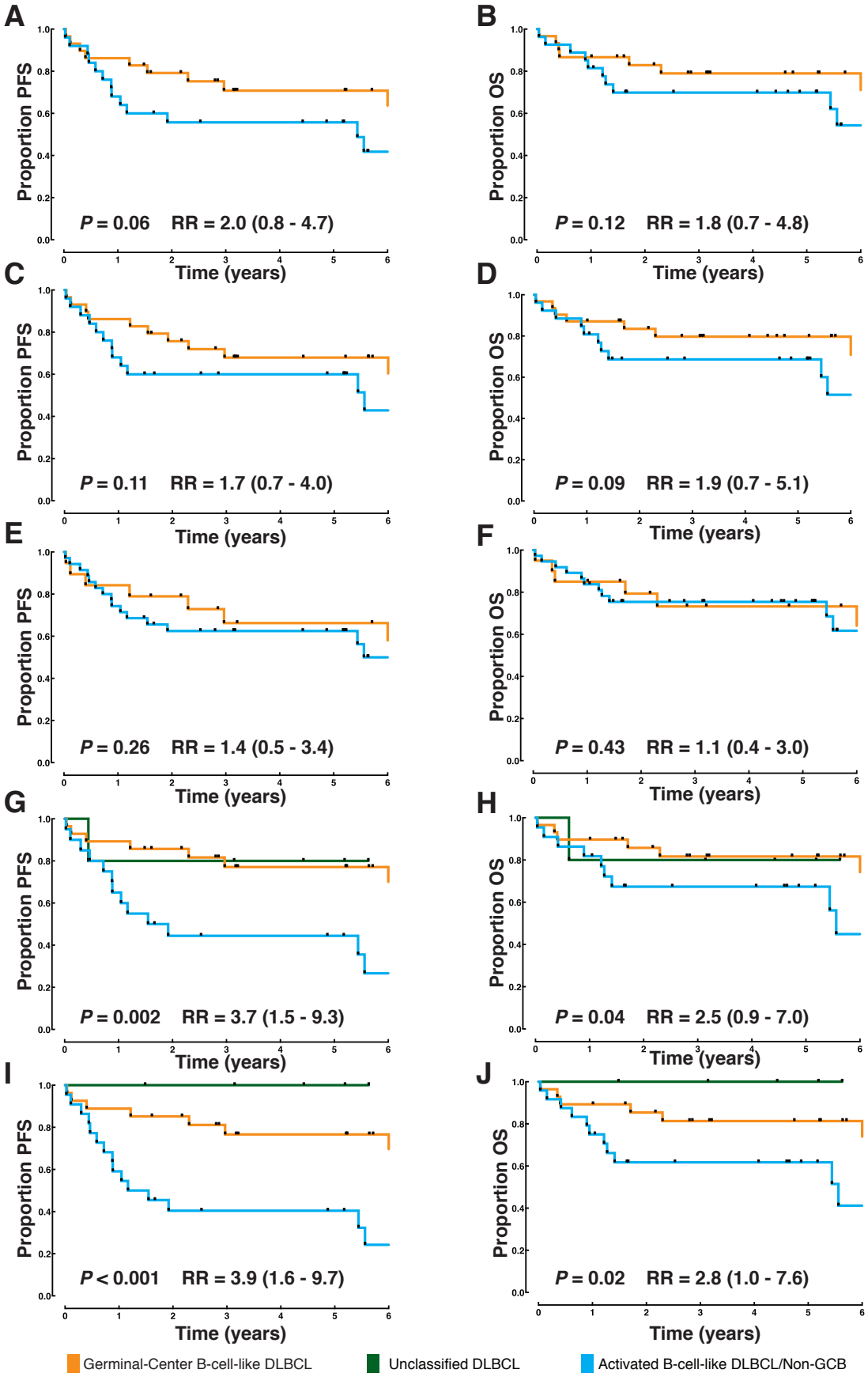

Figure S1: Outcomes in COO groups in the validation cohort. Shown are the Kaplan-Meier curves for the 57 cases of the validation cohort that had valid IHC results. (A), (C), (E), (G) and (I) are for progression free survival. (B), (D), (F), (H) and (J) are for overall survival. (A) and (B) show outcomes for COO assignment using the "Hans" algorithm, (C) and (D) by the "Choi" and (E) and (F) by the "Tally", respectively. For comparison, (G) - (J) are for the Lymph2Cx assay in the same 57 patients using results from the MoCha ((G) and (H)) and CLC ((I) and (J)) sites, respectively. *P* values are from log-rank tests comparing GCB and ABC (or non-GCB) groups. The tests were one-sided in the direction of greater Hazard for ABC. RR: relative risk with 95% confidence interval in brackets.

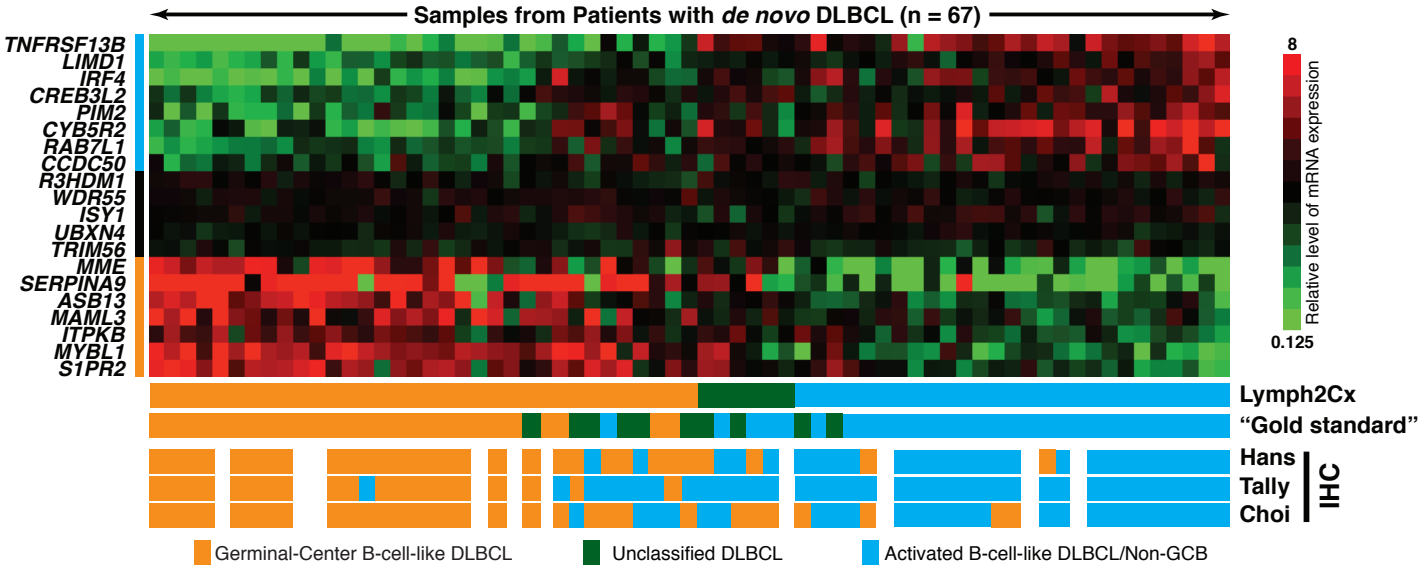

Figure S2: Lymph2Cx model in the independent Validation Cohort at the Centre for Lymphoid Cancer. The 20-gene gene expression COO model (Lymph2Cx) is shown in the form of a heatmap. The patient samples are arrayed vertically, left to right according to ascending Lymph2Cx scores. The genes in the models are shown at the left, with the top 8 genes being over-expressed in ABC-like DLBCL, the middle 5 genes being "housekeeping" genes and the lower 7 genes being over-expressed in GCB-like DLBCL. Below the heatmap, the COO assignment by Lymph2Cx are shown. Below this is the "gold standard" COO from the COO model in Lenz *et al*[1] applied to gene expression profiling using Affymetric U133 Plus 2 arrays on frozen biopsies. Below this are the COO assignments based on immunohistochemistry-based algorithms. Note that 1 of the 68 cases failed QC for the Lymph2Cx assay, while 3 of 60 cases failed to produce interpretable immunohistochemistry results.

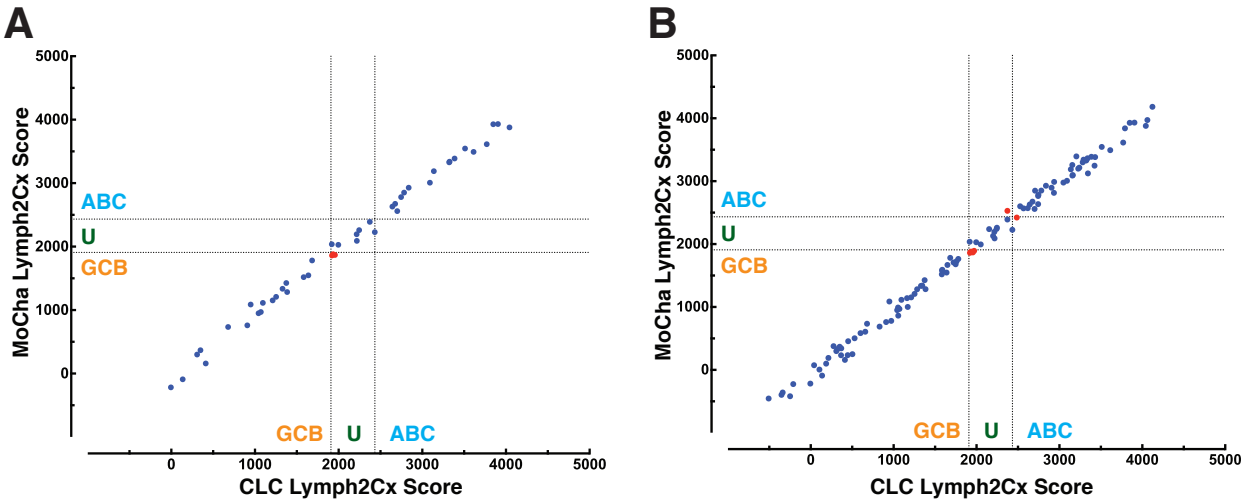

Figure S3: Lymph2Cx scores in the two independent laboratories.
(A) Comparison of the Lymph2Cx scores in the training cohort at the Molecular Characterization Laboratory (MoCha) (FNLCR, Frederick, MD) and the Centre for Lymphoid Cancer (CLC) (BC Cancer Agency, Vancouver, Canada). The dotted lines represent the thresholds between GCB, unclassified and ABC. The R^2 is 0.993 and the slope of the line of best-fit is 1.017.
(B) Comparison of the Lymph2Cx scores in the total cohort (training and validation) between the two sites. The R^2 is 0.995 and the slope of the line of best-fit is 1.015. The red points indicate the cases where there is discordant cell-of-origin assignment at the two sites.

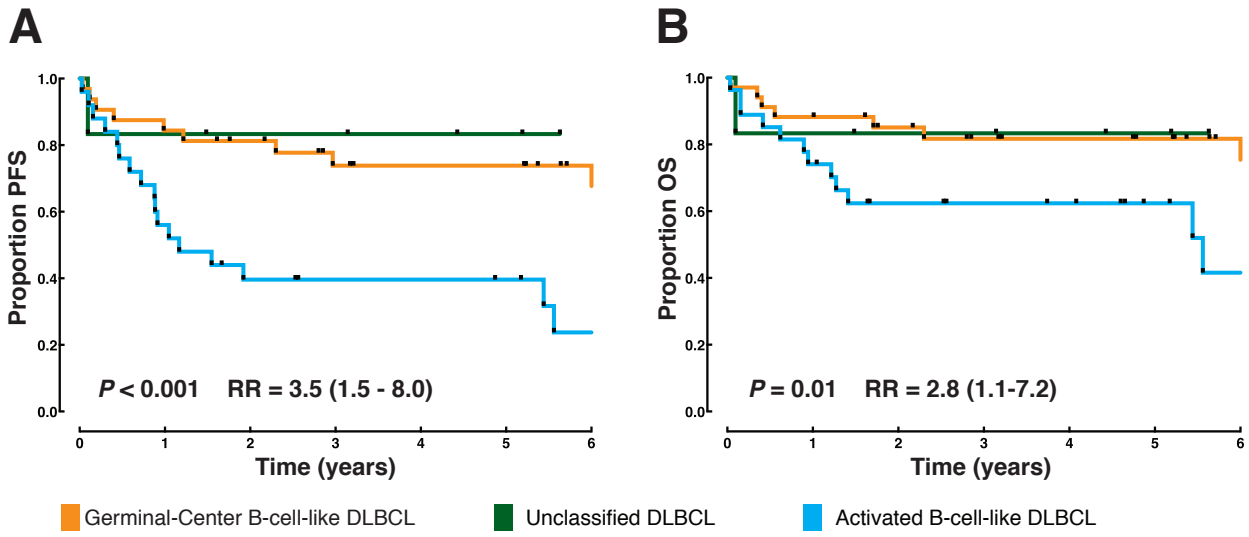

Figure S4: Patient outcomes according to cell-of-origin in the independent validation cohort.
(A) Progression free survival (PFS) in groupings as assigned by the Lymph2Cx at the Centre for Lymphoid Cancer, BC Cancer Agency, Vancouver, Canada. (B) Overall survival (OS) in the same groupings as in (A).
The *P* values are from log-rank tests comparing the ABC and GCB groups. The log-rank tests are one-sided in the direction of greater Hazard for ABC. RR is the relative risk (with 95% confidence interval in brackets) associated with the ABC group compared with the GCB group.

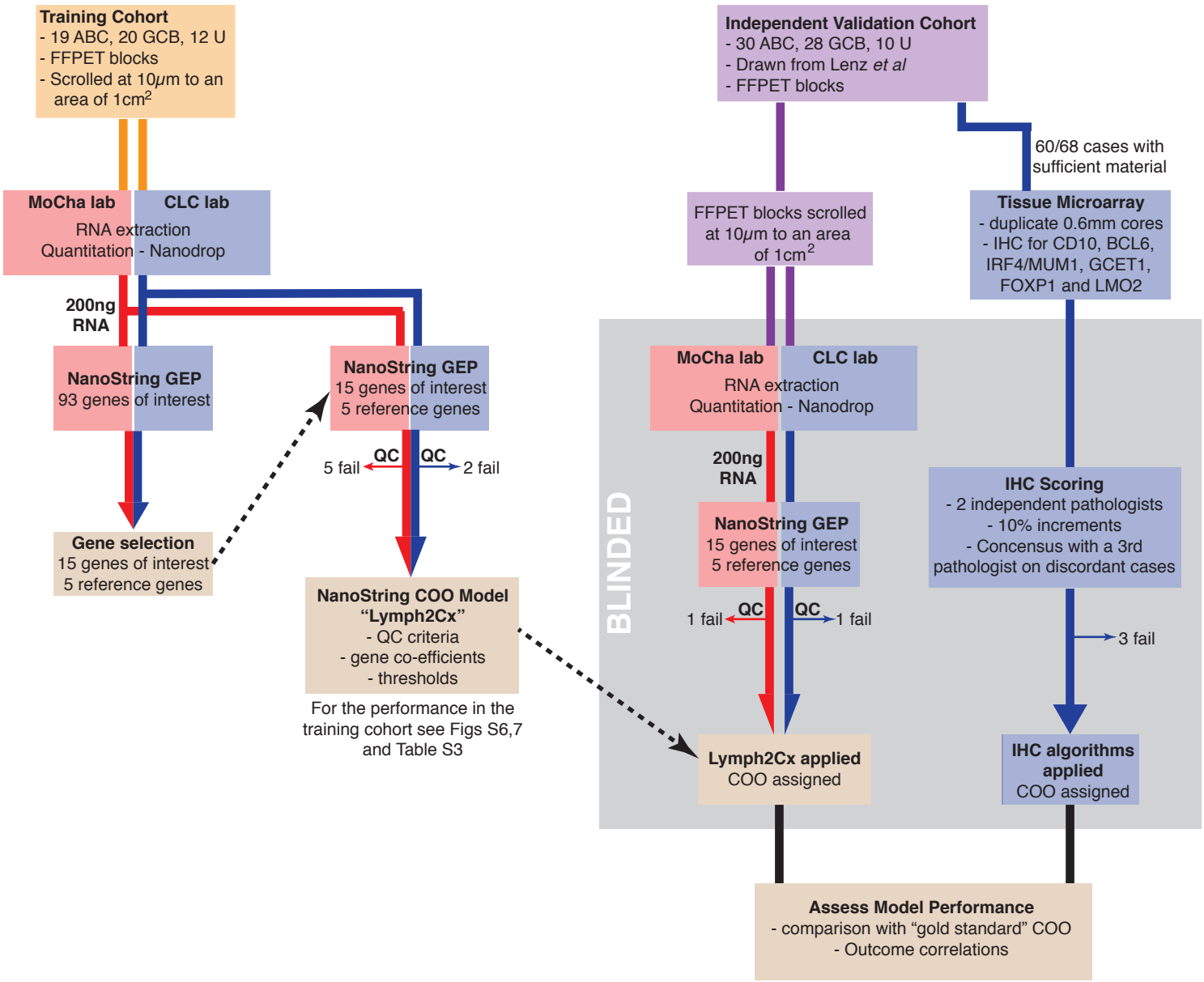

Figure S5: Schematic of the Experimental Design. Shown is the experimental design, with the left hand side showing studies in the training cohort and the right showing studies in the independent validation cohort. Where 2 lines are shown, this indicates that the processes were performed, or occurred, in parallel. Red lines and pink boxes show processes occurring at the Molecular Characterization Laboratory (MoCha), FNLCR, Frederick, MD while blue lines and boxes occurred at the Centre for Lymphoid Cancer (CLC), BC Cancer Agency, Vancouver, Canada. Dotted lines show linkages between processes.

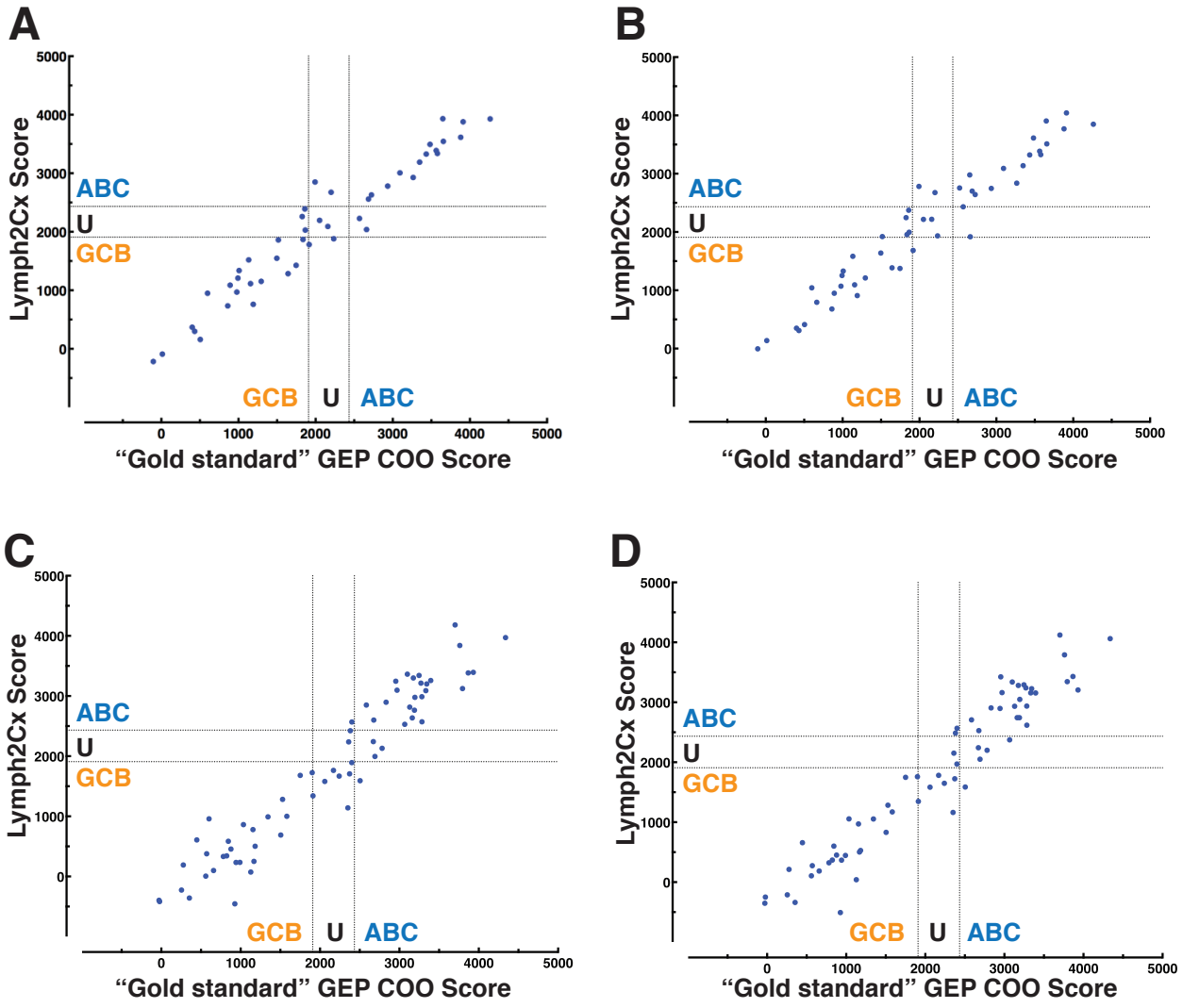

Figure S6: Lymph2Cx scores versus "gold standard" COO scores. In each plot, the score from the NanoString 20-gene model (Lymph2Cx) using RNA from formalin-fixed paraffin-embedded tissue is plotted against the score generated applying the COO model from Lenz *et al*[1] to gene expression generated using Affymetrix U133 Plus 2.0 arrays on matched frozen tissue. The dotted lines represent the threshold scores that separate Germinal Center B-cell-like DLBCL (GCB), unclassified (U) and Activated B-cell-like DLBCL (ABC) . (A) Training cohort extracted and run at the Molecular Characterization Laboratory (FNLCR, Frederick, MD). (B) Training cohort extracted and run at the Centre for Lymphoid Cancer (BC Cancer Agency, Vancouver, Canada). (C) Validation cohort extracted and run at the Molecular Characterization Laboratory. (D) Validation cohort extracted and run at the Centre for Lymphoid Cancer.

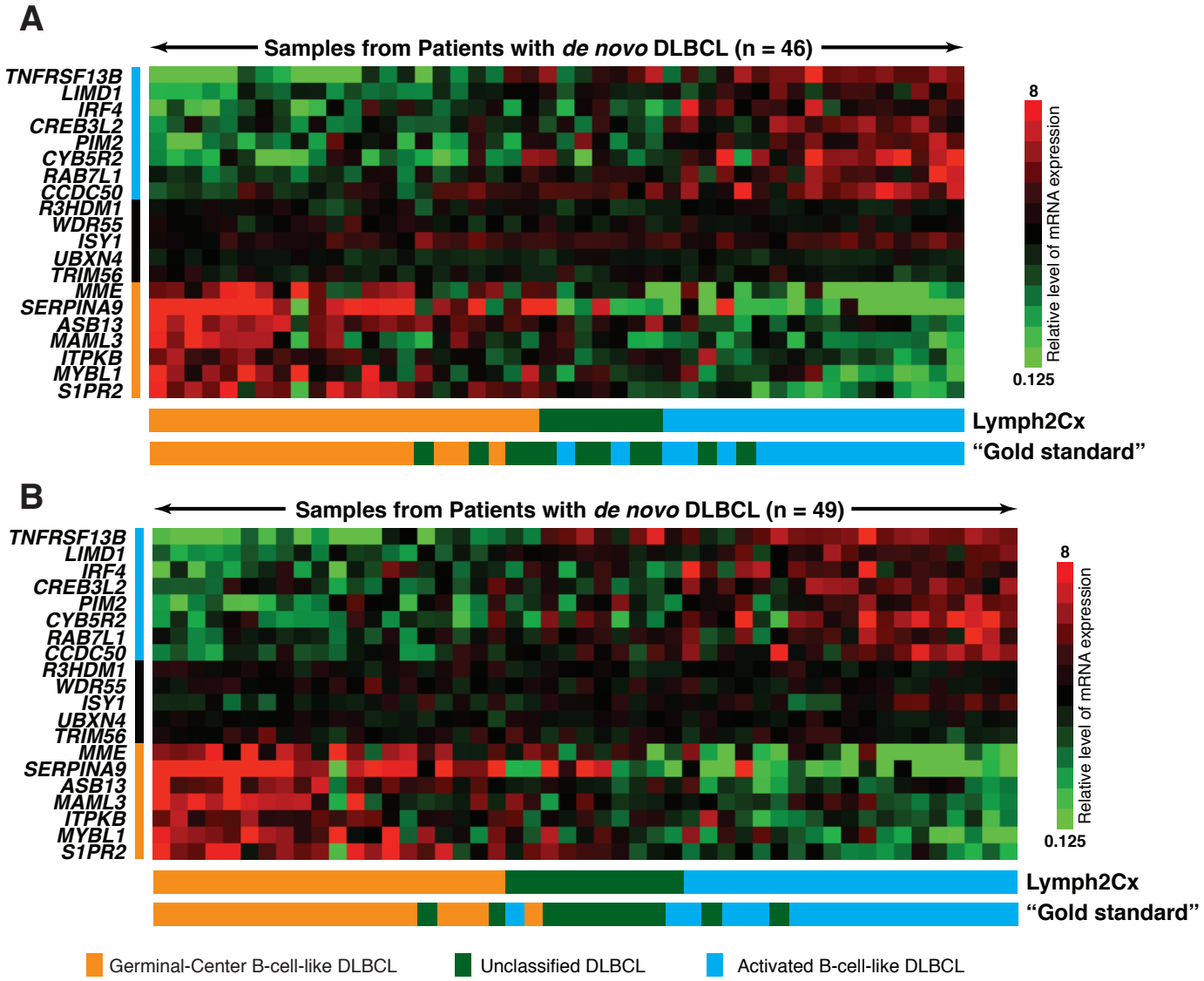

Figure S7: Lymph2Cx model in the Training Cohort. The 20-gene gene expression-based model for COO assignment is shown in the form of heatmaps. The individual samples are arrayed vertically, left to right based on ascending scores from the model. The 20 genes are presented at the left of the heatmap, with the top 8 genes being over-expressed in ABC-like DLBCL, the middle 5 genes being the "housekeeping" genes and the lower 7 genes being over-expressed in GCB-like DLBCL. Below each heatmap the COO assignments of each sample is shown for the Lymph2Cx model and the "gold standard" COO from the COO model in Lenz *et al*[1] applied to gene expression profiling using Affymetric U133 Plus 2 arrays on frozen biopsies. (A) Training cohort extracted and run at the Molecular Characterization Laboratory (FNLCR, Frederick, MD). Note that 5 of the 51 samples in the training cohort failed QC at this site. (B) Training cohort extracted and run at the Centre for Lymphoid Cancer (BC Cancer Agency, Vancouver, Canada). Note that 2 of the 51 samples failed at this site.