Cohort Profile

Cohort Profile: The Lymphoma Specialized Program of Research Excellence (SPORE) Molecular Epidemiology Resource (MER) Cohort Study

James R Cerhan,[1]* Brian K Link,[2] Thomas M Habermann,[3] Matthew J Maurer,[1] Andrew L Feldman,[4] Sergei I Syrbu,[5] Carrie A Thompson,[3] Umar Farooq,[2] Anne J Novak,[3] Susan L Slager,[1] Cristine Allmer,[1] Julianne J Lunde,[1] William R Macon,[4] David J Inwards,[3] Patrick B Johnston,[3] Ivana NM Micallef,[3] Grzegorz S Nowakowski,[3] Stephen M Ansell,[3] Neil E Kay,[3] George J Weiner[2] and Thomas E Witzig[3]

[1]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA, [2]Department of Internal Medicine, University of Iowa Hospitals and Clinics, Iowa City, IA, USA, [3]Department of Medicine, [4]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA and [5]Department of Pathology, University of Iowa Hospitals and Clinics, Iowa City, IA, USA

*Corresponding author. Department of Health Sciences Research, Mayo Clinic, 200 1st Street SW, Rochester, MN 55905, USA. E-mail: cerhan.james@mayo.edu

Why was the cohort set up?

In 2016, an estimated 70 580 people in the USA will have been diagnosed and 20 150 will have died from non-Hodgkin lymphoma (NHL).[1] NHL incidence rates increased over the latter half of the 20th century and only recently stabilized. In parallel, NHL survival rates began improving in the 1990 s with the advent of improved treatment strategies, leading to the current 5-year survival rate of 72%.[2] These trends have led to a growth in the number of NHL survivors, estimated at over 630 000 in the USA in 2013.[3]

In addition to the large and growing population of NHL survivors, NHL presents significant clinical problems for cancer outcomes and survivorship research. First, NHLs comprise a heterogeneous group of diseases that vary in aggressiveness by subtype[4] and vary in outcomes within subtypes by known and unresolved clinical and biological factors.[5] Second, aggressive NHLs require intensive initial therapies which commonly cure patients, but some patients will have treatment sequelae including cardiovascular disease, neuropathies and secondary malignancies.[6] Third, indolent NHLs are rarely cured with standard therapies; but patients often have prolonged survival with the ongoing presence of disease and the effects of serial therapies.[7]

As part of a Lymphoma Specialized Program of Research Excellence (SPORE) programme, the Molecular Epidemiology Resource (MER) was initiated as an observational epidemiology cohort study of prospectively enrolled newly diagnosed lymphoma patients evaluated at the Mayo Clinic (Rochester, MN) and the University of Iowa (Iowa City, IA). The Upper Midwest has some of the

highest lymphoma incidence and mortality rates in the USA.[3] The MER was set up to identify clinical (including comorbid diseases), epidemiological (including lifestyle and other exposures), host (germline genetics, serum/ plasma circulating biomarkers), tumour and treatment factors that impact on multiple outcomes, including event-free, lymphoma-specific and overall survival, new-onset morbidities, patient-reported outcomes (PROs) and general survivorship.

Who is in the cohort?

At both Mayo Clinic Rochester and the University of Iowa, all consecutive cases of lymphoma, including Hodgkin lymphoma (HL) and chronic lymphocytic leukaemia (CLL), who were within 9 months of their initial diagnosis at presentation, a US resident and age 18 years and older, were eligible for enrolment into the MER from 1 September 2002 to 30 June 2015. All participants provided written informed consent, and the cohort protocol was approved by the institutional review boards at the Mayo Clinic and the University of Iowa. Cumulative enrolment was 6972 participants (5256 at Mayo and 1716 at Iowa). Participation rates were 85% at Mayo and 95% at Iowa. The participants were mainly White and from the Upper Midwest. Participant characteristics are shown in Table 1.

How is the cohort followed up?

All participants are systematically contacted every 6 months (± 4 weeks) from the date of original diagnosis for the first 3 years and then annually thereafter. Follow-up data include disease recurrence or progression, new treatments, transformation, new cancer diagnoses and new morbidities; at the 1- and 2-year follow-ups we also obtain data on PROs. We have found that it is most efficient and reliable to conduct follow-ups by mail instead of in-clinic or phone follow-up, which are used as a backup. All reports of disease recurrence, progression, re-treatment or new cancers are validated against medical records. For decedents, we obtain a copy of the death certificate as well as medical records immediately preceding death, in order to review and assign a cause of death by one of the study physicians using a protocol developed for the Eastern Cooperative Oncology Group (ECOG).[8]

An active follow-up protocol allows us to maintain regular contact with participants (including changes in home address and physicians), maintain ARMIs (Authorization to Release Medical Information) for patients being followed outside Mayo or Iowa, efficiently identify and validate new events and obtain follow-up pathology tissue as needed (e.g. at the time of

transformation). We also send an annual newsletter to all participants. However, participants may opt out of the cohort or be followed only through their physician (no direct contact).

As of 1 July 2016 there were 1761 known deaths, 4809 participants in active follow-up, 287 in physician-only follow-up, 27 withdrawals (can use biospecimens), nine withdrawals (must discard biosamples) and 79 lost to clinical follow-up (no ARMI, only followed for mortality). Overall survival (OS) was defined as the time from diagnosis to death due to any cause, and lymphoma-specific survival (LSS) was defined as the time from diagnosis to death due to their cancer. We defined event-free survival (EFS) as the time from diagnosis to disease progression or relapse, re-treatment and/or initiation of second-line therapy, or death due to any cause. Our definition of EFS was chosen over a scan-based progression-free survival (PFS) endpoint, as patients in the MER are managed per treating physician, and not per protocol. Routine clinical care varies widely in scanning and surveillance strategies, and patients in the MER do not have standard disease assessment time points as in a clinical trial. Thus, EFS represents a better clinical endpoint than PFS in the observational cohort setting, and aligns to real-world practice settings and clinical decision points. Outcomes for participants enrolled from 2002 through 2012 ($N = 5445$) and followed through mid-2016 are summarized in Table 2 and in Figure 1a, b. We note that these EFS and OS curves for diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) are similar to those reported by major publications using population-based samples,[9–11] national databases[12,13] and large Phase 3 trials[8,14–16] from the same era.

In Table 3, we compared MER participants enrolled from 2002 through 2012 and aged 20–79 years at enrolment with population-based Surveillance, Epidemiology and End-Results (SEER) data overall and for the state of Iowa (obtained from SEER*Stat 8.3.2.[17]). The participants in the MER who were from Iowa and Minnesota overall were very similar to the overall MER participants in terms of distributions of sex, age, race and NHL subtype, as well as 3-year observed survival (overall and for subgroups). Comparing the MER participants from Iowa and Minnesota with Iowa SEER data on key characteristics shows that the MER has a largely similar distribution (within 5%) on most characteristics in Table 3, with the main exception that the MER somewhat under-represents the age group 70–79 years (21.2% vs 34.4%) and DLBCL (18.9% vs 24.2%). Overall, the 3-year observed survival rate in the MER was 85% (84% for Iowa and Minnesota residents) compared with 77.2% for the state of Iowa and 74.9% for the USA. All differences in observed survival between the various subgroups were <10%, with the largest

Table 1. Baseline demographic characteristics of the Lymphoma SPORE Molecular Epidemiology Resource participants

	DLBCL (N=1348)	FL (N=1121)	CLL/SLL (N=1467)	MCL (N=337)	MZL (N=528)	Other B-cell lymphomas (N=615)	Composite/ discordant (N=299)	HL (N=593)	TCL (N=446)	Other NHL NOS (N=218)	Total (N=6972)
Age at diagnosis, years											
Median	63.0	60.0	63.0	64.0	63.0	63.0	63.0	38.0	58.5	60.0	61.0
Range	(18.0–95.0)	(19.0–93.0)	(23.0–91.0)	(32.0–96.0)	(18.0–92.0)	(18.0–91.0)	(21.0–87.0)	(18.0–89.0)	(18.0–95.0)	(18.0–86.0)	(18.0–96.0)
Age group, years											
18–39	154 (11.4%)	79 (7.0%)	19 (1.3%)	2 (0.6%)	26 (4.9%)	53 (8.6%)	15 (5.0%)	311 (52.4%)	70 (15.7%)	19 (8.7%)	748 (10.7%)
40–49	125 (9.3%)	182 (16.2%)	183 (12.5%)	26 (7.7%)	46 (8.7%)	61 (9.9%)	38 (12.7%)	102 (17.2%)	58 (13.0%)	31 (14.2%)	852 (12.2%)
50–59	279 (20.7%)	278 (24.8%)	372 (25.4%)	88 (26.1%)	136 (25.8%)	133 (21.6%)	65 (21.7%)	56 (9.4%)	100 (22.4%)	58 (26.6%)	1565 (22.4%)
60–69	341 (25.3%)	325 (29.0%)	491 (33.5%)	110 (32.6%)	165 (31.3%)	172 (28.0%)	91 (30.4%)	67 (11.3%)	115 (25.8%)	59 (27.1%)	1936 (27.8%)
70–79	316 (23.4%)	191 (17.0%)	319 (21.7%)	78 (23.1%)	120 (22.7%)	136 (22.1%)	65 (21.7%)	45 (7.6%)	76 (17.0%)	40 (18.3%)	1386 (19.9%)
80 +	133 (9.9%)	66 (5.9%)	83 (5.7%)	33 (9.8%)	35 (6.6%)	60 (9.8%)	25 (8.4%)	12 (2.0%)	27 (6.1%)	11 (5.0%)	485 (7.0%)
Sex											
Female	583 (43.2%)	528 (47.1%)	468 (31.9%)	78 (23.1%)	291 (55.1%)	232 (37.7%)	120 (40.1%)	266 (44.9%)	179 (40.1%)	87 (39.9%)	2832 (40.6%)
Male	765 (56.8%)	593 (52.9%)	999 (68.1%)	259 (76.9%)	237 (44.9%)	383 (62.3%)	179 (59.9%)	327 (55.1%)	267 (59.9%)	131 (60.1%)	4140 (59.4%)
Race											
White	1264 (93.8%)	1043 (93.0%)	1387 (94.5%)	321 (95.3%)	491 (93.0%)	568 (92.4%)	278 (93.0%)	536 (90.4%)	405 (90.8%)	209 (95.9%)	6502 (93.3%)
Non-White	26 (1.9%)	27 (2.4%)	31 (2.1%)	3 (0.9%)	15 (2.8%)	11 (1.8%)	5 (1.7%)	16 (2.7%)	15 (3.4%)	3 (1.4%)	152 (2.2%)
Missing	58 (4.3%)	51 (4.5%)	49 (3.3%)	13 (3.9%)	22 (4.2%)	36 (5.9%)	16 (5.4%)	41 (6.9%)	26 (5.8%)	6 (2.8%)	318 (4.6%)
State of residence at diagnosis											
Iowa	499 (37.0%)	323 (28.8%)	340 (23.2%)	97 (28.8%)	151 (28.6%)	177 (28.8%)	101 (33.8%)	224 (37.8%)	145 (32.5%)	71 (32.6%)	2128 (30.5%)
Minnesota	447 (33.2%)	445 (39.7%)	602 (41.0%)	124 (36.8%)	165 (31.3%)	235 (38.2%)	94 (31.4%)	217 (36.6%)	140 (31.4%)	59 (27.1%)	2528 (36.3%)
Wisconsin	67 (5.0%)	65 (5.8%)	112 (7.6%)	29 (8.6%)	28 (5.3%)	41 (6.7%)	23 (7.7%)	30 (5.1%)	32 (7.2%)	20 (9.2%)	447 (6.4%)
Midwest[a]	270 (20.0%)	216 (19.3%)	291 (19.8%)	66 (19.6%)	149 (28.2%)	126 (20.5%)	63 (21.1%)	104 (17.5%)	95 (21.3%)	50 (22.9%)	1430 (20.5%)
All other	65 (4.8%)	72 (6.4%)	122 (8.3%)	21 (6.2%)	35 (6.6%)	36 (5.9%)	18 (6.0%)	18 (3.0%)	34 (7.6%)	18 (8.3%)	439 (6.3%)

DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; CLL/SLL, chronic lymphocytic leukaemia/small lymphocytic lymphoma; MCL, mantle cell lymphoma; MZL, marginal zone lymphoma; HL, Hodgkin lymphoma; TCL, T-cell lymphoma; NHL, non-Hodgkin lymphoma; NOS, not otherwise specified.

[a]North Dakota, South Dakota, Illinois.

Table 2. Outcomes for the Lymphoma SPORE Molecular Epidemiology Resource participants enrolled 2002 to 2012 and followed through mid-2016

	DLBCL (N=1007)	FL (N=905)	CLL/SLL (N=1197)	MCL (N=257)	MZL (N=420)	Other B-cell lymphomas (N=432)	Composite/discordant (N=253)	HL (N=486)	TCL (N=335)	Other NHL NOS (N=153)	Total (N=5445)
EFS status											
No event	535 (53.1%)	435 (48.1%)	742 (62.0%)	86 (33.5%)	256 (61.0%)	219 (50.7%)	130 (51.4%)	369 (75.9%)	126 (37.6%)	57 (37.3%)	2955 (54.3%)
Event	472 (46.9%)	470 (51.9%)	455 (38.0%)	171 (66.5%)	164 (39.0%)	213 (49.3%)	123 (48.6%)	117 (24.1%)	209 (62.4%)	96 (62.7%)	2490 (45.7%)
Follow-up status											
Alive	635 (63.1%)	741 (81.9%)	884 (73.9%)	131 (51.0%)	352 (83.8%)	287 (66.4%)	180 (71.1%)	419 (86.2%)	175 (52.2%)	67 (43.8%)	3871 (71.1%)
Dead	372 (36.9%)	164 (18.1%)	313 (26.1%)	126 (49.0%)	68 (16.2%)	145 (33.6%)	73 (28.9%)	67 (13.8%)	160 (47.8%)	86 (56.2%)	1574 (28.9%)
Lymphoma-specific survival status											
No event	769 (76.4%)	826 (91.3%)	1068 (89.2%)	170 (66.1%)	400 (95.2%)	349 (80.8%)	208 (82.2%)	452 (93.0%)	225 (67.2%)	96 (62.7%)	4563 (83.8%)
Event	238 (23.6%)	79 (8.7%)	129 (10.8%)	87 (33.9%)	20 (4.8%)	83 (19.2%)	45 (17.8%)	34 (7.0%)	110 (32.8%)	57 (37.3%)	882 (16.2%)
Follow-up years (alive cases)											
N	635	741	884	131	352	287	180	419	175	67	3871
Median	6.1	6.9	6.9	5.9	6.0	5.9	6.1	5.6	5.2	5.6	6.1
Range	(0.0–13.4)	(0.1–13.3)	(0.0–13.5)	(0.4–13.1)	(0.4–13.1)	(0.2–12.9)	(0.1–13.7)	(0.0–13.0)	(0.1–12.9)	(0.4–12.4)	(0.0–13.7)

DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; CLL/SLL, chronic lymphocytic leukaemia/small lymphocytic lymphoma; MCL, mantle cell lymphoma; MZL, marginal zone lymphoma; HL, Hodgkin lymphoma; TCL, T-cell lymphoma; NHL, non-Hodgkin lymphoma; NOS, not otherwise specified.

disparity for the age group 70–79 years (8.9%) and DLBCL (9.1%).

What has been measured?

Pathology review

All cases are reviewed by a haematopathologist and classified based on World Health Organization (WHO) criteria.[18] We collected date of diagnostic biopsy, sampling method (excision, core etc.), grade and selected subtype-specific pathology variables. Up to three WHO subtypes were collected, coded from greatest to least involvement. This allows classification of composite/discordant lymphomas and is an important component of our pathology system, since these lymphomas are not uncommon (∼ 7% of NHL diagnoses) and are not fully captured in the SEER database. This approach allows flexibility in defining cases; for example, in a biopsy sample with FL and areas of DLBCL, the case can be coded as FL for use in aetiology studies (since the low-grade component is considered the primary tumour by SEER rules) and DLBCL for use in prognosis studies (since the DLBCL would be the target of clinical management).

Baseline clinical data abstraction

Clinical data, laboratory values and initial course of therapy at the time of diagnosis were abstracted from primary medical records on all participants; these variables were based on the National Cancer Institute's (NCI's) common data elements (CDEs) for lymphoma.[19] Treatments were either entered as individual agents or as a regimen, allowing extraction of specific agents for analysis. After an initial pilot study, we elected not to collect specific doses, but we do collect numbers of cycles, which allows determination of early discontinuation from standard practice guidelines. Summary descriptive data for selected baseline clinical characteristics, comorbidities and initial treatment are provided in Supplementary Tables 1–3 (available as Supplementary data at *IJE* online), respectively.

Enrolment and follow-up questionnaires

All participants completed a self-administered baseline questionnaire, which includes: race/ethnicity; family history of cancer; history of heart disease, diabetes, hepatitis, shingles, hip fracture, other fracture, osteoporosis, premature menopause, infertility, blood clot, use of blood thinner, organ transplant and autoimmune disorder. Patient-reported outcomes included the Functional Assessment of Cancer Therapy-General (FACT-G),[20] the
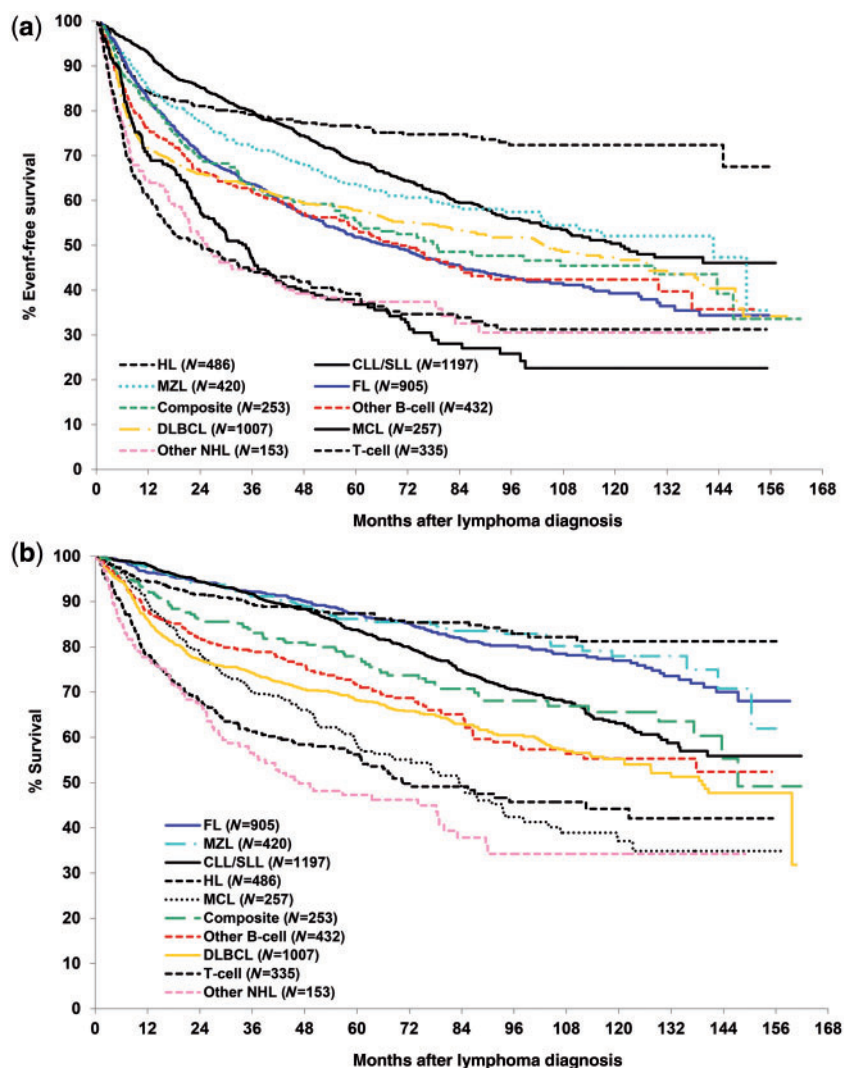
Figure 1. (a) Event-free survival for the Lymphoma SPORE Molecular Epidemiology Resource participants enrolled 2002 to 2012 and followed through mid-2016. (b) Overall survival for the Lymphoma SPORE Molecular Epidemiology Resource participants enrolled 2002 to 2012 and followed through mid-2016.

Linear Analogue Self-Assessment (LASA) quality of life assessment,[21] and performance status. Baseline prevalence of selected comorbidities and median FACT-G scores (normalized 0–100, with a higher score indicating a higher quality of life) are provided in Supplementary Table 2.

At follow-up at 1, 2, 3, 6 and 9 years, we collect the FACT-G and the LASA. The performance of the FACT-G over the first 3 years of patient follow-up in the MER showed that it was valid for monitoring quality of life over time in both aggressive and indolent NHL patients.[22] At 3, 6 and 9 years, we have a more in-depth survivorship mailed questionnaire, with major survey domains shown in Table 4. On a subset of the Mayo participants ($N = 3685$), risk factor data collected as part of a case-control study[23] are available, including detailed medical, reproductive and family history, diet and lifestyle, and farming history, which can also be re-purposed for outcomes studies.

Biorepository

Each participant provided a peripheral blood sample that included two 10-ml EDTA tubes (for plasma and buffy coat for DNA extraction) and two 10-ml tubes for serum. We attempted to obtain samples before the initiation of treatment. For the first EDTA tube, DNA was extracted in batches using an automated salting-out methodology; residual plasma was banked. Genomic DNA was re-suspended in TE buffer, and stored at 4°C using standard protocols. The second EDTA tube was spun, the plasma was removed and aliquoted and the white cell fraction was frozen. The two serum tubes of blood were allowed to clot

Table 3. Comparison of the Molecular Epidemiology Resource (MER; 2002–12, restricted to age 20–79 years) with SEER data

	MER			MER-Minnesota and Iowa			SEER-Iowa			SEER-All		
	N	% dist	3-year observed survival (95% CI)	N	% dist	3-year observed survival (95% CI)	N	% dist	3-year observed survival	N	% dist	3-year observed survival
All	5033	100.0%	85% (84–86)	3299	100.0%	84% (83–85)	9,348	100.0%	77.2%	206,373	100.0%	74.9%
Sex												
Male	2998	59.6%	84% (83–85)	1968	59.7%	83% (81–85)	5321	56.9%	76.0%	118094	57.2%	73.0%
Female	2035	40.4%	86% (84–88)	1331	40.3%	86% (84–88)	4027	43.1%	78.8%	88279	42.8%	77.5%
Age group												
< 40	564	11.2%	90% (88–93)	397	12.0%	91% (88–94)	859	9.2%	91.1%	24194	11.7%	84.4%
40–49	694	13.8%	91% (89–93)	458	13.9%	90% (87–93)	883	9.4%	86.6%	24483	11.9%	81.0%
50–59	1223	24.3%	88% (86–90)	781	23.7%	86% (83–88)	1802	19.3%	82.3%	42633	20.7%	80.3%
60–69	1499	29.8%	84% (82–86)	962	29.2%	81% (81–86)	2587	27.7%	78.1%	55759	27.0%	75.7%
70–79	1053	20.9%	75% (73–78)	701	21.2%	76% (73–80)	3217	34.4%	67.1%	59304	28.7%	63.9%
NHL subtype												
DLBCL	902	17.9%	77% (74–80)	623	18.9%	76% (73–80)	2258	24.2%	66.9%	49280	23.9%	63.9%
FL	849	16.9%	93% (92–95)	588	17.8%	93% (91–95)	1444	15.4%	87.2%	28980	14.0%	86.5%
CLL/SLL	1133	22.5%	92% (91–94)	692	21.0%	93% (91–95)	2253	24.1%	85.0%	40254	19.5%	84.0%
MCL	227	4.5%	72% (67–79)	144	4.4%	73% (66–81)	282	3.0%	64.9%	5617	2.7%	60.1%
MZL	390	7.7%	93% (91–96)	228	6.9%	91% (87–94)	655	7.0%	86.7%	14240	6.9%	87.7%
HL	462	9.2%	91% (88–94)	340	10.3%	91% (88–95)	865	9.3%	88.2%	20899	10.1%	85.6%
TCL	309	6.1%	63% (58–69)	195	5.9%	65% (58–72)	568	6.1%	64.8%	15150	7.3%	64.1%
Race												
White	4821	97.7%	85% (84–86)	3231	97.9%	84% (83–86)	9069	97.0%	77.1%	173441	84.0%	75.6%
Black	112	2.2%	85% (80–92)	68	2.1%	79% (69–89)	122	1.3%	75.0%	17729	8.6%	67.3%

DLBCL, diffuse large B-cell lymphoma; dist, distribution; FL, follicular lymphoma; CLL/SLL, chronic lymphocytic leukaemia/small lymphocytic lymphoma; MCL, mantle cell lymphoma; MZL, marginal zone lymphoma; HL, Hodgkin lymphoma; TCL, T-cell lymphoma; NHL, non-Hodgkin lymphoma.

at room temperature. Clotted blood was sedimented at $800 \times g$ for 10 min and the serum was removed. Plasma was obtained from one EDTA tube that was centrifuged 10 min at $800 \times g$. The supernatant was then removed, centrifuged for an additional 10 min at 800 x g and aliquoted. Serum and plasma were stored at −70°C. Supplementary Table 4 (available as Supplementary data at *IJE* online) shows the available biospecimens, with 90% having extracted DNA and 68% with a buffy coat. Serum (70%) and plasma (78%) were less available, as participants who provided a blood sample off site (returned via overnight delivery service) did not have a serum sample banked.

Paraffin-embedded tumour tissue at Mayo and Iowa is banked in clinical registries (since it is under regulatory control). For cases with tissue blocks outside Mayo or Iowa, pathology reports, slides and tissue are requested and processed through the SPORE Biospecimens Core. Sister tissue microarrays (TMAs) were constructed after review to choose suitable blocks to build a TMA by NHL subtype. Each TMA holds two 1-mm cores from 30 cases plus control samples (e.g. tonsil). To date, TMA construction has been subtype- and project-specific.

What has been found?

Clinical factors

Early disease events
We developed and externally validated (in an independent population) the novel clinical endpoint of event-free survival at 24 months after diagnosis (EFS24) for DLBCL treated with rituximab, cyclophosphamide, doxorubicin, vincristine and prednisolone (R-CHOP).[24] Patients achieving EFS24 have an overall survival equivalent to that of the age- and sex-matched general population ('normal life expectancy'), whereas those not achieving EFS24 have a very poor outcome (and need new therapeutic approaches). We have also developed and externally validated a prognostic model using clinical factors to predict EFS24 (IPI24),[25] and incorporated it into a smartphone app (QxCalculate) for use at the bedside. We have developed and externally validated a similar EFS24 endpoint for immunochemotherapy-treated FL, and more broadly we have introduced EFS12 for FL patients managed with other approaches.[26] These clinical endpoints demonstrate the importance of reassessing prognosis after treatment among survivors, and have direct implications for patient counselling and management, biomarker discovery and trial design.

Table 4. Survey domains and collection time points, Lymphoma SPORE Molecular Epidemiology Resource

Survey domain	Baseline	6 mos	12 mos	18 mos	24 mos	30 mos	3 yrs	4 yrs	5 yrs	6 yrs	7 yrs	8 yrs	9 yrs
Lymphoma status		X	X	X	X	X	X	X	X	X	X	X	X
Cancer status	X	X	X	X	X	X	X	X	X	X	X	X	X
Comorbidity	X	X	X	X	X	X	X	X	X	X	X	X	X
Marital status	X						X			X			X
Employment status	X						X			X			X
Self-perceived health status	X						X			X			X
Weight	X						X			X			X
Performance status	X						X			X			X
FACT-G (QOL)	X		X		X		X			X			X
LASA (QOL)			X		X		X			X			X
Smoking and alcohol	X						X			X			X
Diet	X						X			X			X
Physical Activity	X						X			X			X
STAI							X			X			X
POMS							X						
Cancer screening							X						
CAM and CAM beliefs							X						
Life changes after diagnosis							X						
Stressful life events										X			X
LOT (optimism scale)										X			
Recurrence concerns										X			X
Life changes post cancer										X			X

FACT-G, Functional Assessment of Cancer Therapy-General; QOL, quality of life; LASA, Linear Analogue Self-Assessment; STAI, State-Trait Anxiety Inventory; POMS, Profile of Mood States; CAM, complementary and alternative medicine; LOT, Life Orientation Test; mos, months; yrs, years.

FL transformation

In the pre-rituximab era, transformation from a FL to DLBCL occurred at a rate of ∼ 3%/year and patients with a transformed DLBCL had a median survival of less than 2 years. We reported that the rate of FL transformation to DLBCL in the rituximab era has decreased from 3% to 2%/ year and varied by initial therapy.[27] Patients whose FL transformed to DLBCL had a median survival of 5 years overall, but a poor prognosis was observed in patients with transformation within 18 months of diagnosis or after having received anthracycline therapy, challenging the perception of a universally poor prognosis. The lower rate of transformation was confirmed in the National LymphoCare study.[28]

Statin use at diagnosis

The concurrent use of statins during treatment of patients with DLBCL or FL with rituximab-containing compounds did not affect DLBCL prognosis and was associated with superior FL prognosis,[29] which challenged the clinical significance of a high visibility *in vitro* study that suggested statins inhibit binding of rituximab to CD20.[30] Our results were replicated,[31–34] and statins continue to be widely used by lymphoma patients.

Surveillance scans

Using data from the MER and validated in a Lyon registry, the overwhelming majority of DLBCL relapses were detected outside planned follow-up, with no difference in outcome in patients with DLBCL detected at a scheduled visit compared with relapse detected outside planned follow-up.[35] These data do not support the use of routine surveillance imaging for follow-up of DLBCL, and the findings and conclusions are incorporated into consensus guidelines[36,37] as well as being featured in the American Society of Hematology Choosing Wisely recommendations.

Integrative medicine use

MER patients at 3-year follow-up were surveyed on complementary and alternative medicine (CAM) use: 89% reported ever using any CAM, with 78% using vitamins, 54% alternative therapies and 45% herbal supplements, with more patients using CAM for other health issues than for cancer.[38]

Circulating biomarkers

Vitamin D

Pre-treatment vitamin D insufficiency was associated with an inferior EFS and OS in DLBCL and T-cell lymphoma[39] as well as CLL.[40] As the first identified modifiable prognostic marker in lymphoma, these findings provided the rationale for a Phase III trial (NCT01787409) as well as a

randomized phase II trial (NCT02553447) evaluating outcomes based upon vitamin D replacement strategies.

Immune biomarkers

Elevated serum free light chains (FLC) or an abnormal FLC ratio was associated with an inferior EFS and OS in DLBCL[41] and HL;[42] shorter time to treatment and inferior OS in CLL;[43] and inferior EFS in FL grade 3, mantle cell and other low-grade B-cell lymphomas and peripheral T-cell lymphoma not otherwise specified (PTCL-NOS).[44] DLBCL patients with elevated pre-treatment serum levels of interferon-inducible protein-10 (CXCL10) had inferior EFS and OS even after adjustment for the International Prognostic Index.[45] Elevated pre-treatment serum levels of the pro-inflammatory cytokines IL-2R, IL1-R1 and CXCL9 were associated with inferior EFS in FL.[46] Elevated soluble IL-2Rα levels were associated with EFS in mantle cell lymphoma independent of the mantle cell IPI,[47] whereas elevated levels of serum IL-1RA and sIL-2Rα were independent predictors of EFS in T-cell lymphoma.[48]

Genomics

Tumour genomics

Whole-exome sequencing on 55 primary tumour and paired normal blood samples from MER patients with *de novo* DLBCL identified known and novel recurrent somatic mutations in multiple genes, providing an unbiased view of the mutational landscape in DLBCL.[49] A different set of genomic alterations were identified that distinguished patients who achieve vs fail EFS24, including a 6p21 deletion of *SLC22A16,* a doxorubicin transporter.[50] Comprehensive analysis of FL tumours revealed genetic diversity among newly diagnosed FL patients and found that high tumour complexity and DNA instability may be indicators of more aggressive disease.[51] Novel bioinformatics tools were developed from our studies.[52,53]

Host (germline) genetics

Single nucleotide polymorphisms (SNPs) in genes from the regulators of complement activation (*CFH, CD55, CFHR1, CFHR5, CD46*) at 1q32–q32.1, along with *C9*, were associated with EFS in FL after adjusting for clinical and treatment variables.[54] SNPs in *CXCR5* were associated with EFS in FL patients who were initially observed.[55] In collaboration with the Lymphoma Study Association (LYSA) (French) clinical trials group, we conducted the first genome-wide association study of outcome in R-CHOP-treated DLBCL, identifying 5q23.2 (near *SNX2* and *SNCAIP*) and 6q21 (near *MARCKS* and *HDAC2*) as predictors of EFS and OS.[56]

Tissue-based studies

Using diagnostic tumour specimens from FL patients who later transformed to DLBCL, the presence of PD1[+] T cells and CD14[+] follicular dendritic cells were found to be independent predictors of time to transformation, supporting a role for the tumour microenvironment in the transformation process.[57] *BCL2* somatic mutations at diagnosis were associated with increased risk for transformation and inferior overall survival.[58] IRF4 expression in tumour was associated with poor prognosis in peripheral T-cell lymphomas,[59] suggesting a new targetable pathway. In another study, ALK-negative anaplastic large cell lymphoma was shown to be a genetically heterogeneous disease, with translocations of DUSP22 and TP63 identifying patients with good and poor prognosis, respectively.[60]

Main strengths and limitations

The major strengths of the MER cohort study include: the prospective study design; enrolment of consecutive patients close to their diagnosis (decreasing survival bias); central review and classification of pathology diagnoses according to the WHO criteria; abstraction of key baseline clinical factors that allow computation (and development) of commonly used clinical prognostic indices; and abstraction of initial and subsequent therapies. We have collected and banked a high percentage of biological specimens from participants, including DNA and serum/plasma, and the use of biospecimens is consented to at enrolment for genetic studies and external collaborations. We have access to clinical diagnostic tumour tissue, and have been able to make TMAs for many projects, acknowledging the ongoing challenge of increasing use of core biopsies, limiting tissue availability. Prospective follow-up is tied to the diagnosis date, which generally parallels clinical management (e.g. annual follow-up from diagnosis), facilitating patient and clinician engagement. We collect and validate key outcomes including disease recurrence or progression, retreatment, transformation, new cancers and cause of death. We also collect new morbidities (although these are not validated), PROs and survivorship data. Our study was launched in 2002, shortly after the widespread adoption of the use of rituximab ('rituximab era'), and therefore has high relevance to current clinical practice.

There are also limitations, including that although this is a community-based sample, it is not population-based. We also have limited geographical representation (largely Upper Midwest of the USA), relatively few patients above age 80 and limited ethnic/racial diversity. Obtaining tissue samples or pre-treatment serum/plasma from all patients has not been possible. For the entire cohort, we have

epidemiological data limited to those shown in Table 4, although on the Mayo cases, we can re-purpose data from a case-control study.[23] Sample size is limited for rarer subtypes.

Can I access the MER data? Where can I find out more?

The MER has been under the Iowa/Mayo Lymphoma SPORE collaboration and data sharing plan, which encourages collaboration and use of the resource; external collaborators can contact Drs Cerhan or Link. Whereas use is prioritized for researchers associated with the SPORE Program, all requests are considered by the Executive Committee. Projects must: protect the rights of subjects enrolled in the MER; not interfere with approved projects; advance translational lymphoma research; be financially feasible; provide access to any data generated; and acknowledge the SPORE. Multiple collaborative projects have been approved, and examples of key collaborations have been highlighted above.

Supplementary Data

Supplementary data are available at *IJE* online.

Profile in a nutshell

- The Lymphoma Specialized Program of Research Excellence (SPORE) Molecular Epidemiology Resource (MER) is a prospective cohort study designed to identify clinical, epidemiological, host genetic, biological, tumour and treatment factors that impact on lymphoma outcomes and survivorship.

- Participants were 6972 newly diagnosed lymphoma patients aged 18 years and older, enrolled within 9 months of diagnosis at the Mayo Clinic (Rochester, MN) or the University of Iowa (Iowa City, IA) from 2002 to 2015.

- All participants are contacted every 6 months for the first 3 years after diagnosis and then annually thereafter to ascertain outcomes; disease progression/relapse, re-treatment, transformation and new cancers are validated against medical records. Through July 2016 there have been 1761 deaths, 36 withdrawals and 79 lost to follow-up, with the remainder in active follow-up.

- At enrolment, participants completed a medical history questionnaire, provided a blood specimen (for DNA, plasma and serum) and consented to access to their medical records. Clinical and treatment data were abstracted, and pathology was centrally reviewed. On the Mayo participants, epidemiological data were collected. Patient-reported outcomes and survivorship data were collected at follow-up at 3, 6 and 9 years.

- Access to cohort resources for collaborative research can be requested through the Iowa/Mayo SPORE Data and Biospecimens Access Committee.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016;**66:**7–30.
2. American Cancer Society. *Cancer Facts & Figures 2016.* Atlanta, GA: American Cancer Society, 2016.
3. Howlader N, Noone AM, Krapcho M *et al. SEER Cancer Statistics Review, 1975–2013.* Bethesda, MD: National Cancer Institute, 2016.
4. Flowers CR, Armitage JO. A decade of progress in lymphoma: advances and continuing challenges. *Clin Lymphoma Myeloma Leuk* 2010;**10:**414–23.
5. Shankland KR, Armitage JO, Hancock BW. Non-Hodgkin lymphoma. *Lancet* 2012;**380:**848–57.
6. Flowers CR, Sinha R, Vose JM. Improving outcomes for patients with diffuse large B-cell lymphoma. *CA Cancer J Clin* 2010;**60:**393–408.
7. Kahl BS, Yang DT. Follicular lymphoma: evolving therapeutic strategies. *Blood* 2016;**127:**2055–63.
8. Habermann TM, Weller EA, Morrison VA *et al*. Rituximab-CHOP versus CHOP alone or with maintenance rituximab in older patients with diffuse large B-cell lymphoma. *J Clin Oncol* 2006;**24:**3121–27.
9. Sehn LH, Berry B, Chhanabhai M *et al*. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood* 2007;**109:**1857–61.
10. Lee L, Crump M, Khor S *et al*. Impact of rituximab on treatment outcomes of patients with diffuse large B-cell

lymphoma: a population-based analysis. *Br J Haematol* 2012;**158**:481–88.

11. Keegan TH, McClure LA, Foran JM, Clarke CA. Improvements in survival after follicular lymphoma by race/ethnicity and socio-economic status: a population-based study. *J Clin Oncol* 2009;**27**:3044–51.

12. Friedberg JW, Byrtek M, Link BK *et al*. Effectiveness of first-line management strategies for stage I follicular lymphoma: analysis of the National LymphoCare Study. *J Clin Oncol* 2012;**30**:3368–75.

13. Zhou Z, Sehn LH, Rademaker AW *et al*. An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood* 2014;**123**:837–42.

14. Pfreundschuh M, Muller C, Zeynalova S *et al*. Suboptimal dosing of rituximab in male and female patients with DLBCL. *Blood* 2014;**123**:640–46.

15. Coiffier B, Lepage E, Briere J *et al*. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med* 2002;**346**:235–42.

16. Salles G, Seymour JF, Offner F *et al*. Rituximab maintenance for 2 years in patients with high tumour burden follicular lymphoma responding to rituximab plus chemotherapy (PRIMA): a phase 3, randomised controlled trial. *Lancet* 2011;**377**:42–51.

17. Surveillance, Epidemiology, and End Results (SEER) Program [www.seer.cancer.gov] SEER*Stat Database: Incidence – SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2015 Sub (1973–2013 varying) – Linked to County Attributes – Total U.S., 1969–2014 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2016, based on the November 2015 submission (8 October 2016, date last accessed).

18. Swerdlow SH, Berger F, Pileri SA, Harris NL, Jaffe ES, Stein H. Lymphoplasmacytic lymphoma. In: Swerdlow SH, Campo E, Harris NL *et al*.(eds). *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues* 4th edn. Lyon, France: IARC;2008 (23 December 2002, date last accessed).

19. National Cancer Institute. *CTEP Common Data Elements*.https://wiki.nci.nih.gov/display/caDSR/CTEP+Common+Data+Elements (23 December 2002, date last accessed).

20. Cella DF, Tulsky DS, Gray G *et al*. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol* 1993;**11**:570–79.

21. Locke DE, Decker PA, Sloan JA *et al*. Validation of single-item linear analog scale assessment of quality of life in neuro-oncology patients. *J Pain Symptom Manage* 2007;**34**:628–38.

22. Yost KJ, Thompson CA, Eton DT *et al*. The Functional Assessment of Cancer Therapy – General (FACT-G) is valid for monitoring quality of life in patients with non-Hodgkin lymphoma. *Leuk Lymphoma* 2013;**54**:290–97.

23. Cerhan JR, Fredericksen ZS, Wang AH *et al*. Design and validity of a clinic-based case-control study on the molecular epidemiology of lymphoma. *Int J Mol Epidemiol Genet* 2011;**2**:95–113.

24. Maurer MJ, Ghesquieres H, Jais JP *et al*. Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. *J Clin Oncol* 2014;**32**:1066–73.

25. Maurer MJ, Jais JP, Ghesquieres H *et al*. Personalized risk prediction for event-free survival at 24 months in patients with diffuse large B-cell lymphoma. *Am J Hematol* 2016;**91**:179–84.

26. Maurer MJ, Bachy E, Ghesquieres H *et al*. Early event status informs subsequent outcome in newly diagnosed follicular lymphoma. *Am J Hematol* 2016;**91**:1096–101.

27. Link BK, Maurer MJ, Nowakowski GS *et al*. Rates and outcomes of follicular lymphoma transformation in the immunochemotherapy era: a report from the University of Iowa/MayoClinic Specialized Program of Research Excellence Molecular Epidemiology Resource. *J Clin Oncol* 2013;**31**:3272–78.

28. Wagner-Johnston ND, Link BK, Byrtek M *et al*. Outcomes of transformed follicular lymphoma in the modern era: a report from the National LymphoCare Study (NLCS). *Blood* 2015;**126**:851–57.

29. Nowakowski GS, Maurer MJ, Habermann TM *et al*. Statin use and prognosis in patients with diffuse large B-cell lymphoma and follicular lymphoma in the rituximab era. *J Clin Oncol* 2010;**28**:412–17.

30. Winiarska M, Bil J, Wilczek E *et al*. Statins impair antitumor effects of rituximab by inducing conformational changes of CD20. *PLoS Med* 2008;**5**:e64.

31. Carver JR, Johnson T, Schuster SJ. Rituximab and statins. *J Clin Oncol* 2010;**28**:e611; author reply e612.

32. Ennishi D, Asai H, Maeda Y *et al*. Statin-independent prognosis of patients with diffuse large B-cell lymphoma receiving rituximab plus CHOP therapy. *Ann Oncol* 2010;**21**:1217–21.

33. Samaras P, Heider H, Haile SR *et al*. Concomitant statin use does not impair the clinical outcome of patients with diffuse large B cell lymphoma treated with rituximab-CHOP. *Ann Hematol* 2010;**89**:783–87.

34. Koo YX, Tan DS, Tan IB *et al*. Effect of concomitant statin, metformin, or aspirin on rituximab treatment for diffuse large B-cell lymphoma. *Leuk Lymphoma* 2011;**52**:1509–16.

35. Thompson CA, Ghesquieres H, Maurer MJ *et al*. Utility of routine post-therapy surveillance scans in diffuse large B-cell lymphoma. *J Clin Oncol* 2014;**32**:3506–12.

36. Zelenetz AD, Gordon LI, Wierda WG *et al*. Diffuse Large B-Cell Lymphoma Version 1.2016. *J Natl Compr Cancer Netw* 2016;**14**:196–231.

37. Tilly H, Gomes da Silva M, Vitolo U *et al*. Diffuse large B-cell lymphoma (DLBCL): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2015;**26**(**Suppl 5**):116–25.

38. Rausch Osian S, Leal AD, Allmer C *et al*. Widespread use of complementary and alternative medicine among non-Hodgkin lymphoma survivors. *Leuk Lymphoma* 2015;**56**:434–39.

39. Drake MT, Maurer MJ, Link BK *et al*. Vitamin D insufficiency and prognosis in non-Hodgkin's lymphoma. *J Clin Oncol* 2010;**28**:4191–98.

40. Shanafelt TD, Drake MT, Maurer MJ *et al*. Vitamin D insufficiency and prognosis in chronic lymphocytic leukemia. *Blood* 2011;**117**:1492–98.

41. Maurer MJ, Micallef IN, Cerhan JR *et al*. Elevated serum free light chains are associated with event-free and overall survival in two independent cohorts of patients with diffuse large B-cell lymphoma. *J Clin Oncol* 2011;**29**:1620–26.

42. Thompson CA, Maurer MJ, Cerhan JR et al. Elevated serum free light chains are associated with inferior event free and overall survival in Hodgkin lymphoma. Am J Hematol 2011;86: 998–1000.

43. Maurer MJ, Cerhan JR, Katzmann JA et al. Monoclonal and polyclonal serum free light chains and clinical outcome in chronic lymphocytic leukemia. Blood 2011;118:2821–26.

44. Witzig TE, Maurer MJ, Habermann TM et al. Elevated monoclonal and polyclonal serum immunoglobulin free light chain (FLC) as prognostic factors in B- and T-cell non-Hodgkin lymphoma. Am J Hematol 2014;89:1116–20.

45. Ansell SM, Maurer MJ, Ziesmer SC et al. Elevated pretreatment serum levels of interferon-inducible protein-10 (CXCL10) predict disease relapse and prognosis in diffuse large B-cell lymphoma patients. Am J Hematol 2012;87:865–69.

46. Mir MA, Maurer MJ, Ziesmer SC et al. Elevated serum levels of IL-2R, IL-1RA, and CXCL9 are associated with a poor prognosis in follicular lymphoma. Blood 2015;125:992–98.

47. Sonbol MB, Maurer MJ, Stenson MJ et al. Elevated soluble IL-2Ralpha, IL-8, and MIP-1beta levels are associated with inferior outcome and are independent of MIPI score in patients with mantle cell lymphoma. Am J Hematol 2014;89:E223–27.

48. Gupta M, Stenson M, O'Byrne M et al. Comprehensive serum cytokine analysis identifies IL-1RA and soluble IL-2Ralpha as predictors of event-free survival in T-cell lymphoma. Ann Oncol 2016;27:165–72.

49. Lohr JG, Stojanov P, Lawrence MS et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphona (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci U S A 2012;109:3879–84.

50. Novak AJ, Asmann YW, Maurer MJ et al. Whole-exome analysis reveals novel somatic genomic alterations associated with outcome in immunochemotherapy-treated diffuse large B-cell lymphoma. Blood Cancer J 2015;5:e346.

51. Asmann YW, Maurer MJ, Wang C et al. Genetic diversity of newly diagnosed follicular lymphoma. Blood Cancer J 2014;4:e256.

52. Wang C, Evans JM, Bhagwate AV et al. PatternCNV: a versatile tool for detecting copy number changes from exome sequencing data. Bioinformatics 2014;30:2678–80.

53. Wang C, Davila JI, Baheti S et al. RVboost: RNA-seq variants prioritization using a boosting method. Bioinformatics 2014;30: 3414–16.

54. Charbonneau B, Maurer MJ, Fredericksen ZS et al. Germline variation in complement genes and event-free survival in follicular and diffuse large B-cell lymphoma. Am J Hematol 2012;87: 880–85.

55. Charbonneau B, Wang AH, Maurer MJ et al. CXCR5 polymorphisms in non-Hodgkin lymphoma risk and prognosis. Cancer Immunol Immunother 2013;62:1475–84.

56. Ghesquieres H, Slager SL, Jardin F et al. Genome-wide association study of event-free survival in diffuse large B-cell lymphoma treated with immunochemotherapy. J Clin Oncol 2015;33: 3930–37.

57. Smeltzer JP, Jones JM, Ziesmer SC et al. Pattern of CD14 + follicular dendritic cells and PD1 + T cells independently predicts time to transformation in follicular lymphoma. Clin Cancer Res 2014;20:2862–72.

58. Correia C, Schneider PA, Dai H et al. BCL2 mutations are associated with increased risk of transformation and shortened survival in follicular lymphoma. Blood 2015;125:658–67.

59. Feldman AL, Dogan A, Maurer MJ et al. Expression of interferon regulatory factor-4 (IRF4/MUM1) is associated with inferior overall survival in peripheral T-cell lymphoma (abstract). Blood Cells Mol Dis 2010;116:66.

60. Parilla Castellar ER, Jaffe ES, Said JW et al. ALK-negative anaplastic large cell lymphoma is a genetically heterogeneous disease with widely disparate clinical outcomes. Blood 2014;124:1473–80.