bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Molecular Heterogeneity and Early Metastatic Clone Selection in Testicular Germ Cell Cancer Development

Lambert C.J. Dorssers[1*], Ad J.M. Gillis[1], Hans Stoop[1], Ronald van Marion[1], Marleen M. Nieboer[2], Job van Riet[3,4], Harmen J.G. van de Werken[3,4], J. Wolter Oosterhuis[1], Jeroen de Ridder[2], and Leendert H.J. Looijenga[1*].

[1] Department of Pathology, Wytemaweg 80, 3015 CN, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands.

[2] University Medical Center Utrecht, Universiteitsweg 100, STR 1.305, 3584 CG Utrecht, The Netherlands.

[3] Cancer Computational Biology Center, Wytemaweg 80, 3015 CN, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, Rotterdam, The Netherlands.

[4] Department of Urology, Wytemaweg 80, 3015 CN, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands.

[*] Corresponding authors.

Corresponding authors:

Lambert C.J. Dorssers / Leendert H.J. Looijenga

Dept. of Pathology, JNI, BE 435, Erasmus MC Cancer Institute, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA, Rotterdam, Netherlands.

Phone: 31-10-7044378 /31-10-7044329

Email: l.dorssers@erasmusmc.nl / l.looijenga@erasmusmc.nl

Running title: Nonseminoma Development and Progression

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Background: Testicular germ cell cancer (TGCC), being the most frequent malignancy in young Caucasian males, is initiated from an embryonic germ cell. This study determines intratumor heterogeneity to unravel tumor progression from initiation till metastasis.

Methods: In total 42 purified samples of four treatment-resistant nonseminomatous TGCC (NS) were investigated, including the precursor germ cell neoplasia in situ (GCNIS) and metastatic specimens, using whole genome- and targeted sequencing. Their evolution was reconstructed.

Results: Intratumor molecular heterogeneity did not correspond to the supposed primary tumor histological evolution. Metastases after systemic treatment could be derived from cancer stem cells not identified in the primary cancer. GCNIS mostly lacked the molecular marks of the primary NS and comprised dominant clones that failed to progress. A BRCA-like mutational signature was observed without evidence for direct involvement of *BRCA1* and *BRCA2* genes.

Conclusions: Our data strongly support the hypothesis that NS is initiated by whole genome duplication, followed by chromosome copy number alterations in the cancer stem cell population, and accumulation of low numbers of somatic mutations. These observations of heterogeneity at all stages of tumorigenesis should be considered when treating patients with GCNIS-only disease, or with clinically overt NS.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Background

Malignant germ cell tumors of the adult testis, referred to as type II TGCTs of testicular germ cell cancer (TGCC), are the most frequent cancer in young Caucasian males (1). TGCC are thought to be initiated during early embryogenesis affecting an embryonic germ cell, and become clinically manifest during young adulthood with an annual frequency of approximately 5-12 per 100,000 men in the western world and may require “aggressive” medical treatment. These cancers are clinically and histologically classified into two variants, being seminoma (SE) and nonseminoma (NS). Both arise from a common cancer stem cell, currently referred to as Germ Cell Neoplasia In Situ (GCNIS) (2, 3), which resembles totipotent primordial germ cells (PGC) / gonocytes. Patients with proven GCNIS have a 70% chance of progression to TGCC (both SE and NS) within 7 years. SE consists of a homogeneous population of cells with similarity to GCNIS and PGC/gonocytes. About 50% of the TGCC patients present with a NS that can be composed of different histological elements, embryonal carcinoma (EC), teratoma (TE), yolk sac tumor (YST), and choriocarcinoma, either pure or mixed. The EC is the pluripotent stem cell component of NS, which can mimic normal early embryogenesis including the formation of so called embryonal bodies (EB), and thereby give rise to all differentiated components (3-5).

Although all TGCC, including mature TE, are in principle capable to metastasize, about 80-85% of the SE patients and 55-60% of the NS patients present with localized (stage I) disease. Patients with metastatic TGCC are generally cured by standard treatment regimens involving platinum compounds and additional surgery for residual TE, while only few patients show resistance to treatment (6). So far, detailed studies into the molecular profile of TGCCs and their progression stages were focused on specific genes, like *KIT* (7) and *TP53* (8-10), and on the chromosomal constitution. Analyses have revealed many changes in the (relative) number of individual chromosomes in the different tumor components (5, 11-16). Gain of the

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

short arm of chromosome 12 is a hall mark of (invasive) TGCCs (17), but as yet no causative gene(s) have been identified. Information on driver mutations underlying the development of these NS using exome sequencing is scarce (11-14, 18-20).

In order to unravel the molecular heterogeneity of NS, we extensively investigated four rare cases of primary therapy-resistant NS and performed WGS on the primary cancer and targeted sequencing analyses on 42 enriched histological components, precursor cell populations, and metastatic lesions after treatment (Fig. 1A/B). Focus was on the early events of tumor formation, the molecular heterogeneity within the primary lesion and the retention of molecular markers in the metastatic recurrences. Additionally, data from RNA expression (RNAseq) and copy number alterations (CNA) from high throughput DNA methylation profiling and DEPArray™ /LowPass WGS were interrogated to decipher the evolution of the disease.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and Methods

A brief description is provided here. Further details are provided in the Supplementary Methods.

Patient samples.

NS samples of patients with established intrinsic resistance to standard 1st-line chemotherapy (detailed in Supplementary Methods) were included in this study. Use of tissue samples remaining after diagnosis for scientific reasons was approved by Medical Ethical Committee of the Erasmus MC Rotterdam (The Netherlands), permission 02.981. This included the permission to use the secondary tissue without further consent. Samples were used according to the "Code for Proper Secondary Use of Human Tissue in The Netherlands" developed by the Dutch Federation of Medical Scientific Societies (Version 2002, update 2011).

Omics analyses of patient samples

Purified tumor components (Supplementary Table S1), as defined by an experienced pathologist (JWO), were isolated from frozen tissue slices after staining for alkaline phosphatase enzyme reactivity (21), using PALM micro-dissection (Zeiss). Tumor and paired normal DNA samples were whole genome sequenced (40 times coverage) and analyzed at Complete Genomics Inc. (CG) (Mountain View, CA, USA) using NCBI build 36.3 as human reference genome and pipeline software version 2.0.2.22 (22). Lists of putative somatic DNA variants (SNV) were established from the WGS data as described in the Supplementary Methods. SNV were verified using mutation-specific Q-PCR, targeted sequencing (Supplementary Tables S2&S3), and RNA-seq. Structural variants were evaluated for gene fusions with iFuse (23). Characterization of the mutational signature was done by comparison of the trinucleotide context of each SNV to the established COSMIC signatures using the MutationalPatterns R package (v1.0) (24).

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Targeted sequencing was performed by semiconductor sequencing with the Ion Torrent Personal Genome Machine with supplier's materials and protocols (ThermoFisher Scientific) as previously described (25). Amplicons were designed to cover tumor-specific SNV, structural variants and heterozygote positions (Supplementary Table S3). Median sequencing depth was at least 250 reads. Allele frequencies were established for the heterozygous single nucleotide polymorphisms (SNP) in the matched normal samples present on the amplicons. Details for calling of SNP, SNV and structural variants in the targeted sequencing experiments are provided in the Supplementary Methods. Evolutionary trees of different samples of a specific tumor were drawn based on the LAF and SNV profiles and supported by the TargetClone tool. TargetClone was designed to reconstruct evolutionary trees for multiple samples of a cancer using allele frequencies and SNV (Nieboer MM, Dorssers LCJ, Straver R, Looijenga LHJ, de Ridder J, manuscript submitted and Supplementary Methods).

RNA samples of T6107 and T3209 were rRNA-reduced and Ion Proton sequenced (90 bases, 50 million mapped reads) using the supplier's protocols and reagents (ThermoFisher Scientific). Generation of methylation profiles of primary tumor DNA was performed as previously described (26) or at the Microarray unit of the Genomics and Proteomics Core Facility of the German Cancer Research Center (DKFZ, Heidelberg) strictly adhering to the Illumina EPIC protocols for the T6107-YSTmeta. CNA based on methylation intensities were resolved using the Conumee package (27). DEPArray™ experiments on a T6107 metastatic sample, and GCNIS and YST samples of T618 were performed by Menarini Silicon Biosystems (Castel Maggiore, Italy), essentially as described (28).

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Primary nonseminoma (NS) characteristics

To address tumor heterogeneity and progression, WGS data from four primary chemo-naïve NS were exploited. Comparison of the primary tumors with the matched normal provided a set of 1239 somatic putative DNA variants (SNV) for these cases (Fig. 1B). RNAseq, mutation-specific PCR and targeted sequencing experiments validated 150 out 158 SNV and nine out of 15 structural variants (Fig 1B, Supplementary Tables S4 and S5). The identified mutations causing protein changes have been listed in Table 1. Only four SNV resulted in protein truncation and another 13 were predicted to be damaging (Supplementary Table S6). In addition, detailed information regarding structural variants, lesser allele frequencies (LAF) and chromosome copy number alterations (CNA) were obtained from the WGS and methylation profiling (details in Supplementary Figs. S1 & S2). The trinucleotide profile of single base SNV identified by WGS of the four NS was determined and compared to the established set of COSMIC mutational signatures (29). In all cases (Fig. 1C), signature 3 contributed significantly or was the predominant signature. Analysis of the pooled SNV (from whole exome sequencing) of independent cases of NS (N=18) and SE (N=18) (14), also revealed this signature to be the most prominent (Fig. 1C and further details in Supplementary Fig. S3). This signature 3 is strongly associated with mutations in *BRCA1* or *BRCA2*, genomic deletion and insertion events smaller than 100kb, and a deficiency in homologous recombination repair in breast cancer (30). In the absence of substantial numbers of indels and genomic deletion and insertion events in our WGS data, additional support for recombination repair deficiency is lacking. Furthermore, pathogenic somatic mutations in *BRCA1* or *BRCA2* were not observed in the four included TGCC, although case T1382 carried a predicted, non-pathogenic BRCA1 missense variant (Supplementary Table S6), nor in the cases of the Taylor-Weiner study (14). In addition,

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

promoter hypomethylation and RNAseq reads observed for both genes did not support loss of BRCA function (Supplementary Fig. S4).

Molecular heterogeneity and evolution

For the study of the molecular heterogeneity within these histologically complex primary NS (containing EC, EB, TE and YST components, Fig 1A), matched metastases and precursor lesions, DNA was prepared from various micro- and macro-dissected components (N=42, Fig. 1A/B & Supplementary Table S1). The histological identity was determined by an experienced pathologist, and using direct alkaline phosphatase-staining for EC, EB and GCNIS in frozen tissue (examples in Supplementary Fig. S5) (21). To evaluate the allelic imbalances and the presence of SNV in these enriched specimens, amplicons were designed across the genome containing a tumor-specific SNV and additional heterozygous SNPs (Supplementary Fig. S1 and Supplementary Table S3). For the primary tumor DNA samples, an excellent agreement between the LAF profiles of the WGS and targeted analyses was observed (Supplementary Fig. S6). Analyses of the enriched samples were focused on the LAF of germ line heterozygote SNPs, the read frequencies of the SNV and the presence of specific breakpoints. Furthermore, evolutionary trees based on the LAF and the presence of SNV, and supported by TargetClone, were generated for each case. Results of these analyses will be discussed per case below.

T6107: The majority of the allelic imbalances and SNV found in the primary cancer were present and more easily detected in the enriched malignant histologies due to their increased purity and the removal of contaminating normal cells (example in Fig. 1D). Loss of heterozygosity (LOH) was clearly resolved for chromosome (arm) 4q, 10, 11, 13 and 16q for the purified EC component. The false color plot showed extensive overlap in regional allelic imbalances and SNV amongst the enriched histological components of T6107 and compared to the primary NS (Fig. 2A/B and Supplementary Fig. S7A). LOH on the chromosomes mentioned

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

above was preserved in all histological elements (homogeneously red colored blocks indicating LAF<0.1, Fig. 2A). Sample EC21 displayed additional LOH resulting from copy losses of chromosomes arms 9q and 22q. In the GCNIS preparations (CIS30&FCIS31, Fig. 2A), very little overlap in LAF patterns with the primary NS was observed and the majority of SNV were completely absent (Fig. 2B). The yolk sack tumor metastasis in the lung (YSTmeta) showed minor overlap with the primary NS and its histological components with regard to LAF pattern and presence of SNV. The shared LOH of chromosome arm 22q between EC21 and this metastasis represented independent events based on the different parental alleles retained (Supplementary Fig. S8). Presence of chromosome arm 12p gain (Supplementary Fig. S9), and two SNV (Fig. 2B) in this metastasis demonstrated a shared origin with the primary NS. The copy number profiles of this lung metastasis and a prior retroperitoneal lymph node metastasis with the histology of mature TE displayed many novel alterations, including amplification of the MDM2 region (details in Supplementary Fig. S9). Immunohistochemistry (IHC) and fluorescent in situ hybridization (FISH) analysis confirmed the amplification of MDM2, and targeted sequencing did not reveal TP53 mutation in the lung metastasis (not shown). An evolutionary tree for this case (Fig. 3) was based on the general profiles of the LAF and SNV (Fig. 2A/C), and required two unidentified EC precursors (ECx1 & ECx2) to explain the variance between the primary NS components and the YST lung metastasis. EC21 represented a separate progression line with additional chromosome losses. Furthermore, the early ECx1 precursor containing few aberrations was the founder of the late appearing lung metastasis (YSTmeta) and likely of the mature TE in the lymph node metastasis (Supplementary Fig. S9), which both lacked many of genomic marks of the primary tumor.

T3209: Detection of LOH (chromosomes 4, 14, 15 & 22) was markedly improved for the enriched EC sample, with increased read frequencies for most SNV (Fig. 1E). Absence of specific SNV in this EC sample and present in low read frequencies (10-20%) in the primary tumor indicated clonal variation. Major overlap in LAF patterns (LOH on the above mentioned

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

chromosomes) and most SNV was observed for the enriched histological components and the primary NS (Supplementary Fig. S7B). A single embryonal body (EB23) showed additional regions of LOH (involving chromosomes 1 and 5). The GCNIS preparations essentially lacked allelic imbalances and SNV. The evolutionary tree for this case suggested separate developmental lineages for EB23 and four samples of EC, TE and YST (Fig. 3). The occurrence of very low frequency SNV (<5% of the reads, details Supplementary Fig. S7B) in the GCNIS preparation (which were abundant in the histological components), suggested the presence of a minor population of further progressed GCNIS.

T618: Laser capture of the histological components in the primary NS was not successful due to the presence of excess TE of cartilage differentiation. Purification of YST cells was achieved from FFPE sections using the DEPArray and provided copy number profiles comparable to the primary NS (Supplementary Figs. S1C, S2 & S10). Abundant numbers of GCNIS in the “normal” adjacent parenchyma allowed for the preparation of these cancer stem cells (Supplementary Fig. S5). Subtypes located isolated (CIS29), basal (BCIS28) or floating (FCIS27) and probably reflecting their progression state, were obtained (17). Inspection of the LAF profiles showed increasing allelic imbalances up to LOH for chromosomes 4 and 5 in these GCNIS stages (Fig. 2C/D). In addition, alterations on chromosome arms 11q and 22qter, and 7 SNV were observed in these GCNIS. Copy number analysis of DEPArray purified GCNIS did not reveal gain of chromosome arm 12p (Supplementary Fig, S10). The evolutionary tree of this case (Fig. 3) required the insertion of an unidentified precursor (GCNISx) as the progressed state GCNIS (BCIS28 and FCIS27) lacked 4 SNV (Fig. 2D) and showed loss of alleles on chromosome arms 11q and 22qter, which were retained by the primary tumor (Supplementary Fig. S8). These data strongly suggest that the abundant basal and floating GCNIS adjacent to the tumor mass represent a precursor clone that did not progress to a full malignant state and was not the founder of the primary NS.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

T1382: The enriched samples were all derived from old FFPE tissue blocks and showed more amplicon drop out and noisy LAF data (Supplementary Fig. S7D). In spite of this limitation, the enriched TE and EC histologies showed good overlap for LAF, SNV, and breakpoints with the primary NS. A set of SNV, breakpoints and regions of imbalance appeared preserved in some of the metastases, while novel alterations were also observed (details in Supplementary Fig. S7D). The evolutionary tree based primarily on the SNV required the insertion of a non-identified EC precursor to explain the differences between the primary NS components and the metastases (Fig. 3).

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

Our studies on multiple samples of four cases of therapy-resistant NS provided insight in the complexity of tumorigenesis (Fig. 4). NS carried low numbers of SNV (~0.1 per Mb, Fig. 1B), somewhat lower than reported and comparable to some of the pediatric cancers and spermatocytic tumors (12-14, 20, 31). The total number of SNV with a predicted impact on the encoded protein (3 - 18, Table 1) was similar to other reports (11-13). Overlap of mutated genes belonging to specific pathways was not observed within this small series of TGCC cases and was limited with genes reported for primary untreated TGCC (ASAP1, BRCA1, CCKBR, CROCC, CSE1L, KAT6A, KEAP1, MTOR, MYOM3 and TAF7) (11-14, 20, 32, 33).

NS exhibited clonal heterogeneity which did not correlate with the histological subtypes (Fig. 3). This was expected because of EC being the stem cell component of all differentiated NS elements in the primary NS and different EC precursors providing independent lineages of differentiated cells (5, 34). A typical example is provided by the T3209 EB23 which resembles an early developing embryo derived from a single EC with a different genomic make-up (Supplementary Fig. S7B). Extensive intratumor heterogeneity was also reported for non-small-cell lung cancer (35). Our analyses also showed that metastatic clones can be derived from very early cancer stem cells that are underrepresented or even absent in the primary lesion, and not detected with the current approaches (Fig. 4). Early disseminated cells have been shown to seed metastases in models of breast, pancreatic, bladder and melanoma (36-39), but appeared less predominant in breast cancer patients (40).

The comparison of the primary tumor and the highly enriched tumor histologies revealed that the read frequencies of SNV on the autosomes (even in regions with LOH) hardly ever exceeded the 50% level (examples in Supplementary Figs. S1, S6 and S7), indicating that always a wild type copy of the particular gene was retained within the cancer genome. Similarly, the relative read frequencies of the SNV present in the GCNIS cells were low (< 0.34 for case T618, Fig. 2D) and therefore likely limited to a single allele copy. These results indicate that

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

whole genome duplication preceded the gain of most somatic mutations (35, 40).The single SNV identified in the GCNIS of T6107 was present in high relative read frequencies (up to 50%), suggesting that this mutational event preceded genome duplication (Fig. 2B). Alternatively, this specific variant may represent mosaicism resulting from somatic mutation in early embryonic cells (41, 42). Genome doubling may also underlie development of esophageal cancer following early TP53 inactivation (43). The status of the overrepresentation of the 12p region in the purified cancer stem cells GCNIS remains uncertain, but the observed balanced allele frequencies are in line with absence of 12p gain (Fig. 2A&2C & Supplementary Fig. S8). Furthermore, LowPass WGS on DEPArray purified T618 GCNIS, revealed no gain of 12p and absence of the majority of CNA present in the primary tumor (Supplementary Fig. S10). These results are in agreement with the notion that accumulation of chr12p copies coincides with acquirement of invasive behavior (Fig. 4) (17, 44, 45).

Our data and the lack of recurrent driver mutations support the hypothesis that whole genome duplication is the primary event in NS development (Fig. 4), to be followed by overall net chromosome copy losses (4, 46). Proof for early whole genome duplication may be obtained using digital NGS based on Barcode-In-Genome technology on many individual GCNIS to determine actual chromosome copy numbers (47). Subsequently, gain of 12p copies (which may be dynamic in subclones, details Supplementary Fig. S7), gain of limited numbers of somatic mutations, and additional CNA will trigger the development of the primary NS (Fig. 4). This model is in agreement with the conclusions of Shen et al (20) for the majority of the TGCTs. Only a fraction of *KIT*-mutated SE may have acquired the mutation before whole genome duplication. In order to prevent accumulation of genomic mutations in germ cells, active surveillance and removal of PGCs with an aberrant genome is very efficient (48, 49). All PGCs are completely de-methylated and considered to be prone to aneuploidy (50), but incidence of TGCC is low in the male population. Removal of aberrant GCNIS may require functional TP53 which could be interrupted by gene mutation or by defined miRNAs (51, 52). In agreement with

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

their extreme sensitivity to cisplatin-based therapies, TP53 mutations are extremely rare in primary TGCC indicating no selective pressure (10, 14, 52). The presence of an amplified MDM2 locus in two metastases of case T6107 (Supplementary Fig. S9) may have provided an alternative route for inactivation of TP53 and therapy resistance (10, 32, 53). The BRCA-like base substitution signature of this cancer hints at inefficient homologous recombination repair (54). This BRCA-like signature was not reported in a recent exome sequencing of TGCC (20). There is currently no evidence for the direct involvement of the *BRCA* genes (i.e. absent pathogenic gene mutation and no rearrangement signature (30), and no association of TGCC with familiar *BRCA1/2* mutations), although increased methylation of the *BRCA1* gene promoter was observed in some NS without EC (20). It is tempting to speculate that other components of this homologous recombination repair pathway may be affected and responsible for the specific base substitution signature in the absence of direct involvement of BRCA genes (30, 54). Alternatively, BRCA-related repair functions may be low or turned off intrinsically in the embryonic PGC/gonocytes (55) and during early TGCC development, and thus contribute to the accumulation of this specific pattern of base substitutions. Therefore, TGCC development may be the result of the properties of the PGC (de-methylated and reduced repair pathway), which allows the tetraploid GCNIS to evade apoptosis and survive till puberty and subsequently progress to malignancy due to genomic destabilization (56). The (epi)-genetic triggers for destabilization of the tetraploid genome and initiation of the development of malignant clones remain as yet largely unknown, although the induction of mitogenic signaling by testosterone may contribute.

The observed heterogeneity in the primary tumor, metastases and precursor lesions of nonseminoma may impact on clinical decisions and treatment strategies. The occurrence of metastatic tumors with little molecular overlap with the primary lesion indicates that treatment of therapy-resistant recurrences should be targeted at their molecular properties according to the concepts of personalized medicine. The identification of abundant cancer stem cells GCNIS

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

which did not contribute to the development of the NS further complicates the clinical advice to patients with GCNIS only disease (57). In view of its tendency to progress in 70% of the patients within 7 years, novel predictive markers for the GCNIS progression are needed. The BRCA-like mutational signature in TGCC indicative for reduced homologous recombination repair, may be suggestive for combined use of PARP inhibitors and platinum-based therapy (20, 58) but requires additional support. Further studies into NS and SE cases, with GCNIS adjacent to the primary tumor and different histologies or localization, are needed to identify reliable markers for GCNIS progression, to unravel the critical steps for malignancy and therapy resistance, and to decipher the origin of the BRCA-like mutational signature of TGCC.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Additional Information.

Ethics approval.

Use of tissue samples remaining after diagnosis for scientific reasons was approved by Medical Ethical Committee of the Erasmus MC Rotterdam (The Netherlands), permission 02.981. This included the permission to use the secondary tissue without further consent. Samples were used according to the "Code for Proper Secondary Use of Human Tissue in The Netherlands" developed by the Dutch Federation of Medical Scientific Societies.

Availability of data and materials

WGS, ENA PRJEB20644, accession numbers ERX2100523 – 530; targeted sequencing, ENA PRJEB20644, accession numbers: ERX2019898-958; RNA-seq, ArrayExpress, accession number: E-MTAB-5746; DNA methylation, GEO (GSE58538, GSM1413103 to GSM1413106) or ArrayExpress (E-MTAB-5842, sample T6107-YSTmeta).

Conflict of interest

The authors declare no potential conflicts of interest.

Funding

Financial support: Whole genome sequencing was provided by Complete Genomics, Inc. Mountain View, CA 94043, USA, and DEPArray experiments were performed by Menarini Silicon Biosystems S.p.a. (Castel Maggiore, Italy).

Authors' contribution

LD and LL: design of experiments and writing the manuscript. AG, HS, RM, LD: execution of experiments, JvR, HW, MN, JdR & LD for bioinformatics analyses, JWO: pathology review. All authors approved the manuscript.

Acknowledgements

We greatly acknowledge Lieke Dons and Dragana Bertovic for assistance in sample purification. We thank Remko Hersmus for performing FISH experiments; Daphne Heijsman, Arne S. IJpma

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and Peter van de Spek for help with WGS data; Saskia D. Hiltemann, and Andrew P. Stubbs for support with Galaxy and analysis of structural variants (IFuse); Ans van den Ouweland for functional evaluation of SNV; Erik-Jan Dubbink for targeted sequencing options, and Alex Nigg for the plug-in to enhance FISH images. We thank the microarray unit of the DKFZ Genomics and Proteomics Core Facility for providing the Illumina Human Methylation arrays and related services, and Volker Hovestadt and colleagues for help with the copy number analyses of methylation intensity data. We also acknowledge the Erasmus MC Cancer Computational Biology Center for giving access to their IT-infrastructure and software that was used for computations and/or data analysis in this study. This study was supported in part by a WGS grant from Complete Genomics, Inc. Mountain View, CA 94043, USA. We appreciate the support from Rosanna Lanzellotto, Alberto Ferrarini, Bechir Boughaba, Raimo Tanzi and Francesca Fontana from Menarini Silicon Biosystems S.p.a. Castel Maggiore (Italy) for the DEPArray experiments.

Supplementary information is available at the British Journal of Cancer's website

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Ulbright TM, Amim MB, Balzer B, Berney DM, Epstein JI, Guo C, et al. Germ cell tumours. In: H. M, Humphrey PA, Ulbright TM, Reuter VE, editors. WHO Classification of Tumours of the Urinary System and Male Genital Organs. 4th ed. Lyon: IARC Press; 2016. p. 189–226.

2. Berney DM, Looijenga LH, Idrees M, Oosterhuis JW, Rajpert-De Meyts E, Ulbright TM, et al. Germ cell neoplasia in situ (GCNIS): evolution of the current nomenclature for testicular pre-invasive germ cell malignancy. Histopathology. 2016;69(1):7-10. Epub 2016/02/27.

3. Spiller CM, Bowles J. Germ cell neoplasia in situ: The precursor cell for invasive germ cell tumors of the testis. Int J Biochem Cell Biol. 2017;86:22-5. Epub 2017/03/16.

4. Oosterhuis JW, Looijenga LH. Testicular germ-cell tumours in a broader perspective. Nat Rev Cancer. 2005;5(3):210-22. Epub 2005/03/02.

5. Killian JK, Dorssers LC, Trabert B, Gillis AJ, Cook MB, Wang Y, et al. Imprints and DPPA3 are bypassed during pluripotency- and differentiation-coupled methylation reprogramming in testicular germ cell tumors. Genome Res. 2016;26(11):1490-504. Epub 2016/11/03.

6. Beyer J, Albers P, Altena R, Aparicio J, Bokemeyer C, Busch J, et al. Maintaining success, reducing treatment burden, focusing on survivorship: highlights from the third European consensus conference on diagnosis and treatment of germ-cell cancer. Ann Oncol. 2013;24(4):878-88. Epub 2012/11/16.

7. McIntyre A, Summersgill B, Grygalewicz B, Gillis AJ, Stoop J, van Gurp RJ, et al. Amplification and overexpression of the KIT gene is associated with progression in the

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

seminoma subtype of testicular germ cell tumors of adolescents and adults. Cancer Res. 2005;65(18):8085-9. Epub 2005/09/17.

8. Chaganti RS, Houldsworth J. Genetics and biology of adult human male germ cell tumors. Cancer Res. 2000;60(6):1475-82. Epub 2000/04/05.

9. Oosterhuis JW, Peeters SH, Smit VT, Stoop H, Looijenga LH, Elzevier HW, et al. Patient with two secondary somatic-type malignancies in a late recurrence of a testicular non-seminoma: illustration of potential and flaw of the cancer stem cell therapy concept. Int J Dev Biol. 2013;57(2-4):153-7. Epub 2013/06/21.

10. Kersemaekers AM, Mayer F, Molier M, van Weeren PC, Oosterhuis JW, Bokemeyer C, et al. Role of P53 and MDM2 in treatment response of human germ cell tumors. J Clin Oncol. 2002;20(6):1551-61. Epub 2002/03/16.

11. Brabrand S, Johannessen B, Axcrona U, Kraggerud SM, Berg KG, Bakken AC, et al. Exome sequencing of bilateral testicular germ cell tumors suggests independent development lineages. Neoplasia. 2015;17(2):167-74. Epub 2015/03/10.

12. Cutcutache I, Suzuki Y, Tan IB, Ramgopal S, Zhang S, Ramnarayanan K, et al. Exome-wide Sequencing Shows Low Mutation Rates and Identifies Novel Mutated Genes in Seminomas. Eur Urol. 2015;68(1):77-83. Epub 2015/01/19.

13. Litchfield K, Summersgill B, Yost S, Sultana R, Labreche K, Dudakia D, et al. Whole-exome sequencing reveals the mutational spectrum of testicular germ cell tumours. Nature communications. 2015;6:5973. Epub 2015/01/23.

14. Taylor-Weiner A, Zack T, O'Donnell E, Guerriero JL, Bernard B, Reddy A, et al. Genomic evolution and chemoresistance in germ-cell tumours. Nature. 2016;540(7631):114-8. Epub 2016/12/03.

15. de Jong B, Oosterhuis JW, Castedo SM, Vos A, te Meerman GJ. Pathogenesis of adult testicular germ cell tumors. A cytogenetic model. Cancer genetics and cytogenetics. 1990;48(2):143-67. Epub 1990/09/01.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

16. Oosterhuis JW, Castedo SM, de Jong B, Cornelisse CJ, Dam A, Sleijfer DT, et al. Ploidy of primary germ cell tumors of the testis. Pathogenetic and clinical relevance. Lab Invest. 1989;60(1):14-21. Epub 1989/01/01.

17. Rosenberg C, Van Gurp RJ, Geelen E, Oosterhuis JW, Looijenga LH. Overrepresentation of the short arm of chromosome 12 is related to invasive growth of human testicular seminomas and nonseminomas. Oncogene. 2000;19(51):5858-62. Epub 2000/12/29.

18. Bilen MA, Hess KR, Campbell MT, Wang J, Broaddus RR, Karam JA, et al. Intratumoral heterogeneity and chemoresistance in nonseminomatous germ cell tumor of the testis. Oncotarget. 2016;7(52):86280-9. Epub 2016/11/20.

19. Ichimura K, Fukushima S, Totoki Y, Matsushita Y, Otsuka A, Tomiyama A, et al. Recurrent neomorphic mutations of MTOR in central nervous system and testicular germ cell tumors may be targeted for therapy. Acta neuropathologica. 2016;131(6):889-901.

20. Shen H, Shih J, Hollern DP, Wang L, Bowlby R, Tickoo SK, et al. Integrated Molecular Characterization of Testicular Germ Cell Tumors. Cell reports. 2018;23(11):3392-406. Epub 2018/06/14.

21. Stoop H, Kirkels W, Dohle GR, Gillis AJ, den Bakker MA, Biermann K, et al. Diagnosis of testicular carcinoma in situ '(intratubular and microinvasive)' seminoma and embryonal carcinoma using direct enzymatic alkaline phosphatase reactivity on frozen histological sections. Histopathology. 2011;58(3):440-6. Epub 2011/02/18.

22. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010;327(5961):78-81. Epub 2009/11/07.

23. Hiltemann S, McClellan EA, van Nijnatten J, Horsman S, Palli I, Teles Alves I, et al. iFUSE: integrated fusion gene explorer. Bioinformatics. 2013;29(13):1700-1. Epub 2013/05/11.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

24. Blokzijl F, Janssen R, Van Boxtel R, Cuppen E. MutationalPatterns: an integrative R package for studying patterns in base substitution catalogues. http://biorxivorg/content/early/2016/10/17/071761. 2016.

25. Dubbink HJ, Atmodimedjo PN, Kros JM, French PJ, Sanson M, Idbaih A, et al. Molecular classification of anaplastic oligodendroglioma using next-generation sequencing: a report of the prospective randomized EORTC Brain Tumor Group 26951 phase III trial. Neuro Oncol. 2016;18(3):388-400. Epub 2015/09/12.

26. Rijlaarsdam MA, Tax DM, Gillis AJ, Dorssers LC, Koestler DC, de Ridder J, et al. Genome wide DNA methylation profiles provide clues to the origin and pathogenesis of germ cell tumors. PLoS One. 2015;10(4):e0122146. Epub 2015/04/11.

27. Hovestadt V, Zapatka M. conumee: Enhanced copy-number variation analysis using Illumina DNA methylation arrays. R package version 1.6.0. http://bioconductororg/packages/conumee/. 2015.

28. Bolognesi C, Forcato C, Buson G, Fontana F, Mangano C, Doffini A, et al. Digital Sorting of Pure Cell Populations Enables Unambiguous Genetic Analysis of Heterogeneous Formalin-Fixed Paraffin-Embedded Tumors by Next Generation Sequencing. Scientific reports. 2016;6:20944. Epub 2016/02/13.

29. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21. Epub 2013/08/16.

30. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. 2016;534(7605):47-54. Epub 2016/05/03.

31. Giannoulatou E, Maher GJ, Ding Z, Gillis AJM, Dorssers LCJ, Hoischen A, et al. Whole-genome sequencing of spermatocytic tumors provides insights into the mutational processes operating in the male germline. PLoS ONE. 2017;12(5):e0178169. Epub 2017/05/26.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

32. Bagrodia A, Lee BH, Lee W, Cha EK, Sfakianos JP, Iyer G, et al. Genetic Determinants of Cisplatin Resistance in Patients With Advanced Germ Cell Tumors. J Clin Oncol. 2016(34):4000-7. Epub 2016/09/21.

33. TCGA-TGCT. Testicular Germ Cell Cancer (TCGA, Provisional). 2016:Accessed 28 Sep 2016.

34. Stevens LC. Experimental production of testicular teratomas in mice of strains 129, A/He, and their F1 hybrids. J Natl Cancer Inst. 1970;44(4):923-9. Epub 1970/04/01.

35. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. N Engl J Med. 2017;376(22):2109-21. Epub 2017/04/27.

36. Hosseini H, Obradovic MM, Hoffmann M, Harper KL, Sosa MS, Werner-Klein M, et al. Early dissemination seeds metastasis in breast cancer. Nature. 2016;50:552-8. Epub 2016/12/16.

37. Rhim AD, Mirek ET, Aiello NM, Maitra A, Bailey JM, McAllister F, et al. EMT and dissemination precede pancreatic tumor formation. Cell. 2012;148(1-2):349-61. Epub 2012/01/24.

38. Eyles J, Puaux AL, Wang X, Toh B, Prakash C, Hong M, et al. Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma. J Clin Invest. 2010;120(6):2030-9. Epub 2010/05/27.

39. Shin K, Lim A, Odegaard JI, Honeycutt JD, Kawano S, Hsieh MH, et al. Cellular origin of bladder neoplasia and tissue dynamics of its progression to invasive carcinoma. Nat Cell Biol. 2014;16(5):469-78. Epub 2014/04/22.

40. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, et al. Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell. 2017;32(2):169-84 e7. Epub 2017/08/16.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

41. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. Science. 2015;350(6256):94-8. Epub 2015/10/03.

42. Ju YS, Martincorena I, Gerstung M, Petljak M, Alexandrov LB, Rahbari R, et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. Nature. 2017;543(7647):714-8. Epub 2017/03/23.

43. Stachler MD, Taylor-Weiner A, Peng S, McKenna A, Agoston AT, Odze RD, et al. Paired exome analysis of Barrett's esophagus and adenocarcinoma. Nat Genet. 2015;47(9):1047-55. Epub 2015/07/21.

44. Summersgill B, Osin P, Lu YJ, Huddart R, Shipley J. Chromosomal imbalances associated with carcinoma in situ and associated testicular germ cell tumours of adolescents and adults. Br J Cancer. 2001;85(2):213-20. Epub 2001/07/20.

45. Ottesen AM, Skakkebaek NE, Lundsteen C, Leffers H, Larsen J, Rajpert-De Meyts E. High-resolution comparative genomic hybridization detects extra chromosome arm 12p material in most cases of carcinoma in situ adjacent to overt germ cell tumors, but not before the invasive tumor development. Genes Chromosomes Cancer. 2003;38(2):117-25. Epub 2003/08/27.

46. van Echten J, Oosterhuis JW, Looijenga LH, van de Pol M, Wiersema J, te Meerman GJ, et al. No recurrent structural abnormalities apart from i(12p) in primary germ cell tumors of the adult testis. Genes Chromosomes Cancer. 1995;14(2):133-44. Epub 1995/10/01.

47. Xi L, Belyaev A, Spurgeon S, Wang X, Gong H, Aboukhalil R, et al. New library construction method for single-cell genomes. PLoS ONE. 2017;12(7):e0181163. Epub 2017/07/21.

48. Salmina K, Jankevics E, Huna A, Perminov D, Radovica I, Klymenko T, et al. Up-regulation of the embryonic self-renewal network through reversible polyploidy in irradiated p53-mutant tumour cells. Exp Cell Res. 2010;316(13):2099-112. Epub 2010/05/12.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

49. Fotovati A, Nakayama K, Nakayama KI. Impaired germ cell development due to compromised cell cycle progression in Skp2-deficient mice. Cell division. 2006;1:4. Epub 2006/06/09.

50. Hoffman RM. Is DNA methylation the new guardian of the genome? Molecular cytogenetics. 2017;10:11. Epub 2017/04/12.

51. Lane DP. Cancer. p53, guardian of the genome. Nature. 1992;358(6381):15-6. Epub 1992/07/02.

52. Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell. 2006;124(6):1169-81. Epub 2006/03/28.

53. Bauer S, Muhlenberg T, Leahy M, Hoiczyk M, Gauler T, Schuler M, et al. Therapeutic potential of Mdm2 inhibition in malignant germ cell tumours. Eur Urol. 2010;57(4):679-87. Epub 2009/06/30.

54. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet. 2017;49(10):1476-86. Epub 2017/08/22.

55. Ahmed EA, van der Vaart A, Barten A, Kal HB, Chen J, Lou Z, et al. Differences in DNA double strand breaks repair in male germ cell types: lessons learned from a differential expression of Mdc1 and 53BP1. DNA repair. 2007;6(9):1243-54. Epub 2007/03/23.

56. van der Zwan YG, Biermann K, Wolffenbuttel KP, Cools M, Looijenga LH. Gonadal maldevelopment as risk factor for germ cell cancer: towards a clinical decision model. Eur Urol. 2015;67(4):692-701. Epub 2014/09/23.

57. Elzinga-Tinke JE, Dohle GR, Looijenga LH. Etiology and early pathogenesis of malignant testicular germ cell tumors: towards possibilities for preinvasive diagnosis. Asian J Androl. 2015;17(3):381-93. Epub 2015/03/21.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

58. Petljak M, Alexandrov LB. Understanding mutagenesis through delineation of mutational signatures in human cancer. Carcinogenesis. 2016;37(6):531-40. Epub 2016/05/22.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Legends

Figure 1. Dissection of nonseminoma. **A**. NS may consist of multiple histological elements (examples case T1382, TE and EC; case T618, GCNIS, OCT3/4 stained nuclei), which were each enriched by PALM-assisted micro dissection. The primary NS was subjected to WGS and the purified components to targeted sequencing using Ion Torrent technology. **B**. The results of the WGS are shown per case. Complete Genomics (CG) SNV were retrieved from CG output files, and putative candidates were selected after visual inspection of the reads. SNV validation was performed by mutation-specific PCR, targeted DNA sequencing and/or RNA sequencing (Supplementary Table S4). Structural variants (SV) were confirmed for selected cases using targeted sequencing. Mutations occurring near exon boundaries were evaluated for potential effect on splicing using Alamut Software. Details are provided in Supplementary Tables S4-S6. The bottom panel shows the number of purified histological components isolated and analyzed for these cases. PBL = peripheral blood leukocytes; NAP = nonmalignant adjacent parenchyma. **C**. The mutational profiles of these four NS were compared to the 30 COSMIC signatures described. The mutational signatures contributing significantly (>5%) are presented. In addition, the SNV of NS and SE cases from the Taylor-Weiner study (*) (14) were pooled for tumor subtype and similarly analyzed. The number of SNV used for the analysis is indicated below each bar. **D/E**. Profiles of lesser allele frequencies (LAF) of heterozygote SNP and relative read frequencies of SNV derived from targeted sequencing of primary NS (red) and dissected EC components (green) are shown. SNV are presented as filled symbols (NS: yellow diamonds; EC: green squares). Positions are provided on chromosomes scaled according to size.

Figure 2. False color plots of allelic imbalances, SNV and structural variants in primary TGCC (NS) and various purified tumor components of two cases (T6107 (panels A/B) and T618 (panels C/D)). Data from (tumor) samples analyzed more than once were averaged. **Top panels A and C:** Each line represents the amplicon averaged LAF of heterozygote SNPs using

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the LAF color scheme in panel E (LAF).Blue indicates heterozygosity and red refers to LOH. Ordering is based on chromosomal position (indicated on the right). For comparison, the 100kb interval WGS LAF (CG-LAF) is also shown. In addition, the WGS relative read coverage (CGrelCov) data of the primary NS are provided for the specific chromosomal regions using the color scheme (Rel Cov) in panel E. Missing data are white. Sample types have been marked by colored dashed lines (GCNIS: orange, histological components: red and metastases: purple). (*) FFPE tissue blocks derived DNA samples. **Bottom panels B and D:** Occurrence of tumor-specific SNV and structural variants in the different tumor samples (grey indicates absence, black indicates more than 3% of the reads carrying the variant, missing data in white). Tumor-specific structural variants are indicated with green arrow heads. Genes with a mutation resulting in amino acid change (T618: only those observed in GCNIS) have been indicated (see Table 1). A red arrow marks a SNV present in all samples (except PBL) from case T6107 (chr19:56131557), residing in a 2kb region between two Zn-finger genes. The number at the bottom indicates the months after surgery of the primary tumor for the removal of the metastasis. **E:** Color keys for the different categories. Sample information and targeted sequence data details are provided in Supplementary Table S1 and Table S8.

Figure 3. Evolution of nonseminoma. Evolutionary trees for the different histological specimens for all cases are shown on this developmental model. Coloring of boxes is according to sample type: GCNIS (yellow), precursor EC (light orange), primary tumor (dark orange) and metastasis (purple). The order and grouping of the samples is based on the similarities in allele and SNV profiles (as shown in Fig. 2 and Supplementary Fig. S7) and supported by TargetClone (Materials and Methods section), and implementing the biological constraint for NS development that the differentiated components (TE and YST) originated from an EC-type precursor. Samples with comparable profiles are grouped together. In specific cases of samples with partially non-overlapping features (i.e. SNV and/or LOH) without an immediate precursor, a non-

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

identified precursor (GCNISx or ECx) was inserted at the branch point in the tree (T6107, T618). Similarly, to comply with the evidence that an EC is the precursor of TE and YST, a non-identified precursor ECx was introduced (T6107, T1382). Specific gains of LOH or SNV compared to their immediate precursor are indicated. *) indicates SNV present at very low read frequencies, suggesting polyclonality for GCNIS or the presence of a minor subclone in some of the primary tumor components of T3209. For case T1382, a minimal tree is presented with ordering of samples based on SNV only due to sequencing noise.

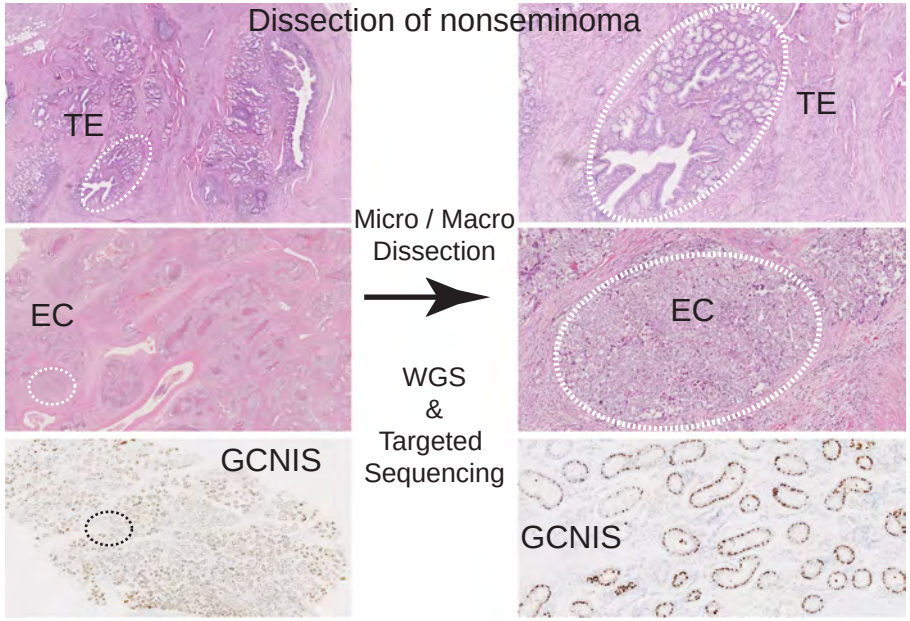
Figure 4. Nonseminoma evolution. A model of the different steps in NS development from normal PGC/gonocytes to metastasized NS is shown. The model is based on the available literature regarding early genome duplication, acquisition of extra copies of the short arm of chromosome 12, the pluripotent capacity of the EC (4), and the results from this study. Following initial whole genome doubling, during puberty chromosome loss may be the predominant way to change the copy numbers in the formation of GCNIS cancer stem cell (yellow). GCNIS represent a polyclonal mixture of cells, some may remain dormant and others may progress to malignancy. The gain of copies of 12p ($12p^{++}$) is a hallmark of the precursor with invasive potential (light orange). Further losses and gains of chromosomes or chromosome fragments may contribute to the formation of the primary tumor with its distinct histological components (orange) and the typical metastases (purple). Somatic mutation appears to be limited and occurring at later stages. The metastases may also originate from precursor EC not detected in the primary tumor.

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tumor	Symbol	ACC	Mutation Type	NT	AA	Presence of SNV in		
						GCNIS	Histology	Meta
T1382	EIF5B	NM_015904	missense	c.541A>C	p.N181H		1/1	7/7
	TTN	NM_003319	missense	c.5558T>A	p.I1853N		2/2	5/8
	LPCAT1	NM_024830	missense	c.1129G>A	p.E377K		2/2	7/7
	LPCAT1	NM_024830	missense	c.517G>C	p.V173L		2/2	8/8
	CHRNA2	NM_000742	missense	c.688G>A	p.E230K		2/2	1/8
	KAT6A	NM_006766	missense	c.158T>C	p.L53S		2/2	4/4
	SYT7	NM_001252065	missense	c.655G>A	p.G219S		1/1	n.d.
	BICD1	NM_001003398	missense	c.490C>T	p.R164W		2/2	1/8
	FREM2	NM_207361	missense	c.7882C>T	p.R2628W		2/2	5/6
	POLR3K	NM_016310	stopgain	c.49G>T	p.G17X		2/2	8/8
	SNX20	NM_182854	missense	c.911T>A	p.I304N		2/2	8/8
	SETD6	NM_024860	missense	c.1318A>G	p.I440V		2/2	0/6
	CA5A	NM_001739	missense	c.721G>A	p.E241K		2/2	8/8
	BRCA1	NM_007300	missense	c.745A>G	p.T249A		2/2	4/7
	TMEM147	NM_032635	missense	c.104G>A	p.C35Y		2/2	1/1
	SLCO4A1	NM_016354	missense	c.175C>T	p.L59F		2/2	8/8
T3209	MTOR	NM_004958	missense	c.3011G>C	p.C1004S	0/1	13/13	
	NRXN1	NM_001135659	missense	c.634G>T	p.G212C	n.d.	6/6	
	TAF7	NM_005642	nonfrs del	c.738_740del	p.I247del			
	PDLIM7	NM_005451	missense	c.422C>T	p.P141L	n.d.	11/12	
	CCKBR	NM_176875	missense	c.1240C>G	p.R414G	n.d.	2/9	
	RNF219	NM_024546	missense	c.1940A>T	p.Q647L	n.d.	5/5	
	GCOM1	NM_001285900	missense	c.1027G>C	p.E343Q	0/1	3/13	
	ABCC3	NM_003786	acceptor	c.2415-5C>T	Unknown	0/1	13/13	
	ZNFX1	NM_021035	frs del	c.3592_3593del	p.S1198AfsX71			
	SHOX	NM_006883	missense	c.310G>A	p.V104M	n.d.	8/8	
T6107	CPSF3	NM_016207	missense	c.762T>G	p.D254E	0/2	6/6	0/1
	UACA	NM_018003	missense	c.2272G>C	p.D758H	n.d.	3/3	0/1
	DNAJB1	NM_006145	missense	c.200A>G	p.Y67C			
T618	CROCC	NM_014675	nonfrs sub	c.244_246	p.Q82del	3/3		
	ARHGEF10L	NM_018125	acceptor	c.-43-3C>T	Unknown	0/3		
	MYOM3	NM_152372	nonfrs sub	c.3688_3689delinsAG	p.Q1230R	0/3		
	TTF2	NM_003594	missense	c.634C>A	p.H212N	0/1		
	KIAA0226	NM_014687	missense	c.1963C>T	p.H655Y	3/3		
	SMARCAD1	NM_001254949	missense	c.226A>G	p.N76D			
	ASAP1	NM_001247996	missense	c.933G>C	p.Q311H	0/2		
	DNAJB5	NM_001135005	frs sub	c.1077_1093G	p.Nfs??	0/3		
	ZDHHC6	NM_022494	missense	c.950G>A	p.R317H			
	CLCF1	NM_001166212	missense	c.468A>T	p.E156D	0/3		
	NDUFV1	NM_007103	acceptor	c.511-4G>A	Unknown	1/2		
	LRRC10	NM_201550	missense	c.125G>A	p.R42H	0/3		
	PIAS1	NM_016166	missense	c.1712A>C	p.D571A	0/3		
	KEAP1	NM_203500	missense	c.610C>T	p.R204W	3/3		
	CSRP2BP	NM_020536	stopgain	c.412G>T	p.E138X	1/2		
	CSE1L	NM_001316	donor	c.1619+4A>T	unknown	3/3		
	APOL3	NM_145640	missense	c.1066C>G	p.L356V	0/3		
	CFP	NM_002621	missense	c.1285G>T	p.V429L	3/3		

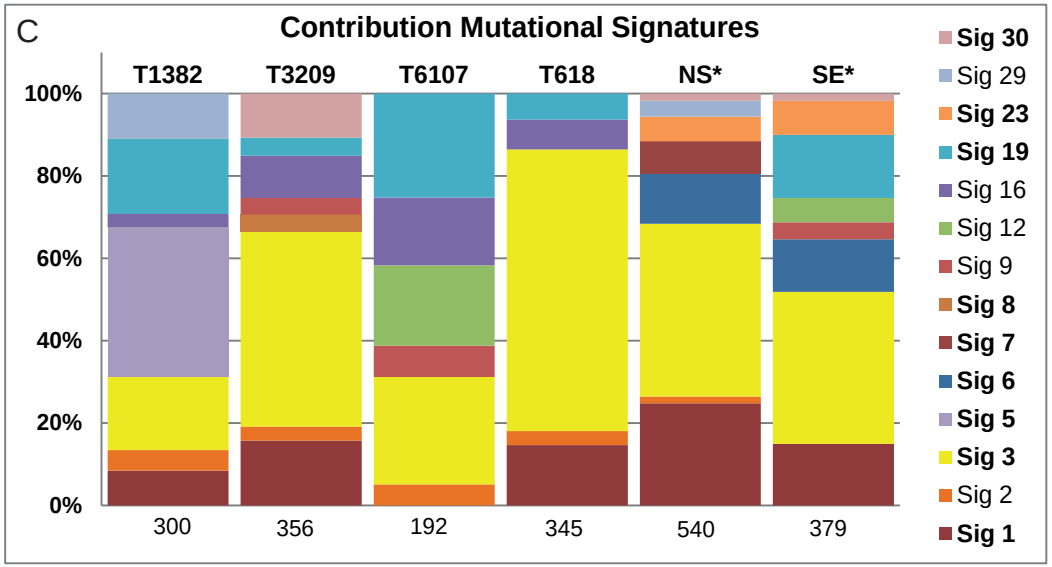
Table 1. Somatic mutation derived protein variants. Somatic mutation of genes leading to (putative) protein changes per case. The mutation type, RNA nucleotide (NT) and amino acid (AA) changes are indicated. The last three columns provide the occurrence of the DNA mutation (samples with mutation / samples successfully sequenced) in the different enriched sample types (GCNIS, histologies of the primary tumor, and metastases). n.d. = not detected, indicating that the sequencing of the specific target was not successful.

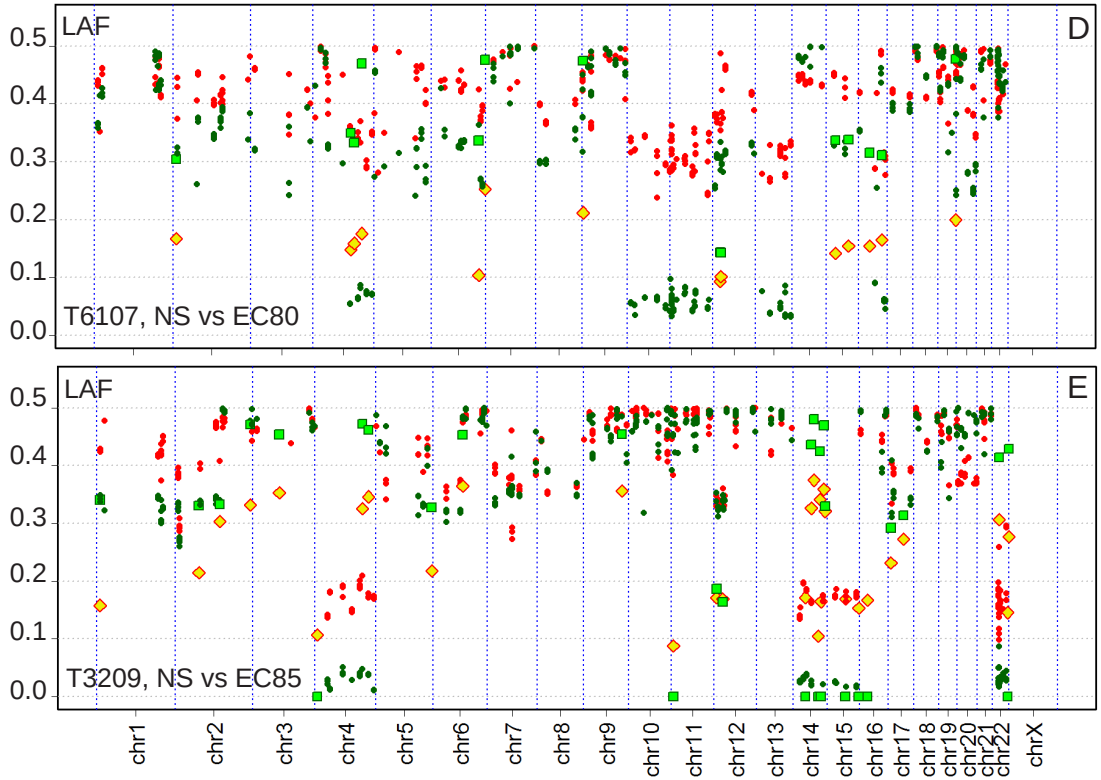
A

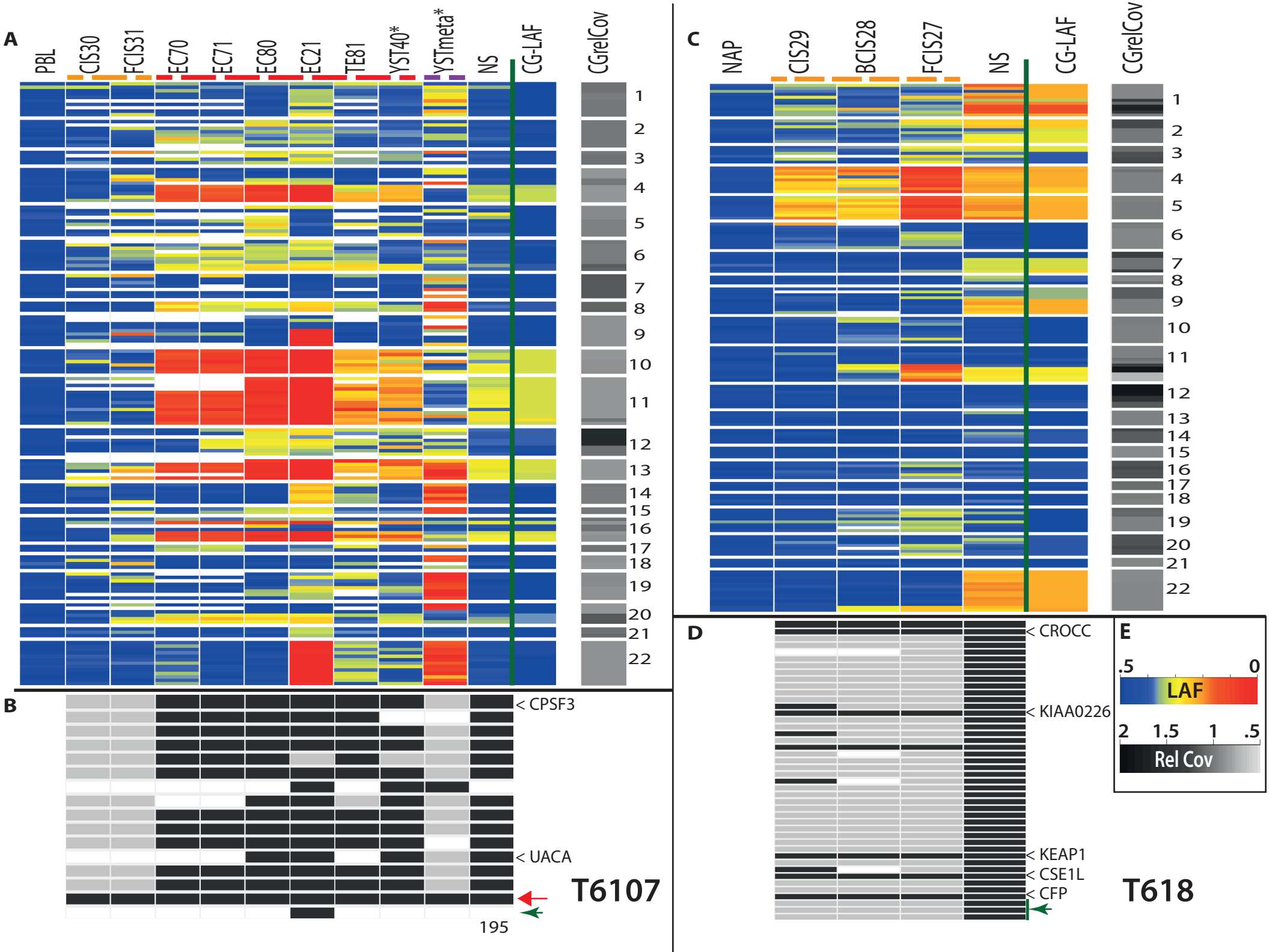


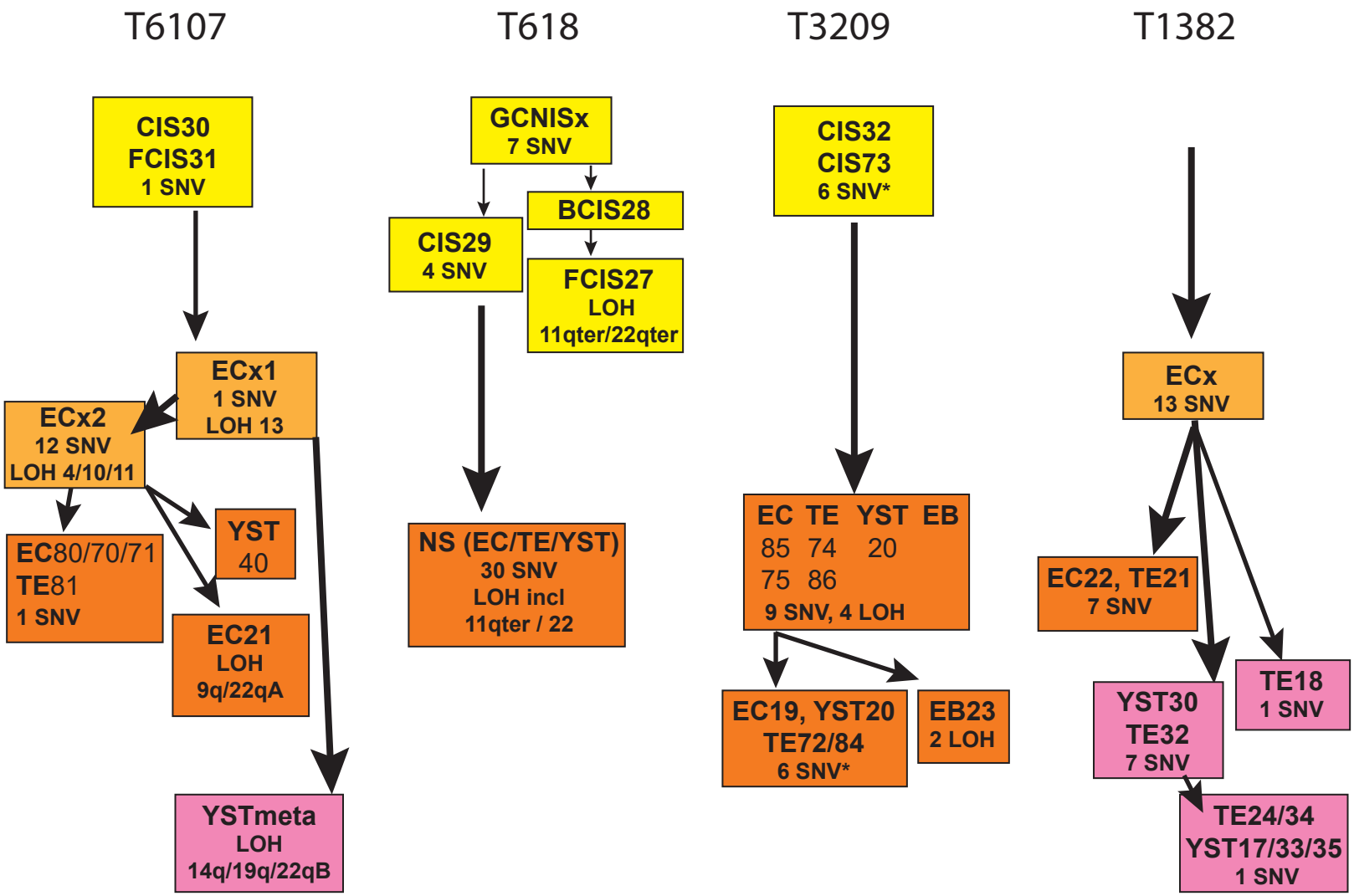
B

Case	T3209	T6107	T618	T1382
	NS / PBL	NS / PBL	NS / NAP	NS / PBL
Whole Genome Sequencing				
CG SNV	558	635	455	494
Putative SNV	374	197	365	303
Validated SNV / tested	55 / 56	24 / 25	40 / 45	31 / 32
CG SV	33	9	26	24
Validated SV / tested	2 / 2	1 / 1	4 / 8	2 / 4
Splicing mutation	0 / 4	1 / 2	3 / 4	0 / 1
Protein change	12	3	14	16
Micro- or macro dissection of specific components				
GCNIS	2	2	3	
EC	3	4		1
EB	7			
TE	4	1		5
YST	3	3		4
Normal			1	



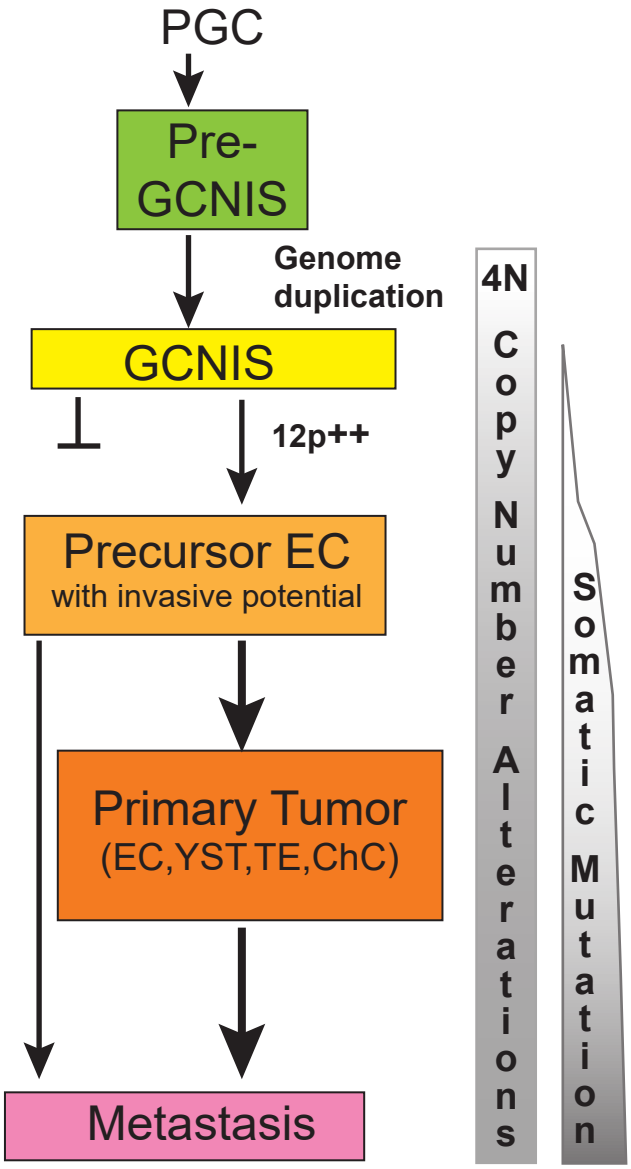



A
PBL
CIS30
FCIS31
EC70
EC71
EC80
EC21
TE81
YST40*
YSTmeta*
NS
CG-LAF
CGrelCov
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
B
< CPSF3
< UACA
195
T6107
C
NAP
CIS29
BCIS28
FCIS27
NS
CG-LAF
CGrelCov
D
< CROCC
< KIAA0226
< KEAP1
< CSE1L
< CFP
T618
E
.5
0
LAF
2
1.5
1
.5
Rel Cov



Nonseminoma cases

bioRxiv preprint doi: https://doi.org/10.1101/385807; this version posted August 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



Nonseminoma Model