Development of a Unified DNA and RNA NGS Panel for Thyroid Cancer

OVERVIEW

The Informagen Next Generation Sequencing (NGS) Thyroid Test offers a model for the development of a multi-categorical assay for simultaneous interrogation of RNA and DNA markers. The Informagen test enables the in-depth molecular characterization of thyroid cancer fine-needle aspirate (FNA) and formalin-fixed paraffin-embedded (FFPE) specimens through NGS analysis of hotspot regions across a 104-gene targeted DNA/RNA-Seq panel. This multi-panel assay, compatible with downstream Illumina sequencing chemistry, is composed of 68 amplicons covering clinically actionable mutations in *BRAF, KRAS, NRAS, HRAS, PIK3CA* and 15 additional genes, 52 RNA translocations and aberrant isoforms, and 40 mRNA disease-relevant expression markers.

High-value diagnostic and theranostic DNA/RNA targets for the Informagen NGS Thyroid Test were selected based on established guidelines, published reports, the Cancer Genome Atlas (TCGA) and internal discovery efforts (Haynes et al, 2015, in preparation). A performance assessment was completed for the Informagen test to support verification and anticipate validation of a thyroid cancer marker assay to replace and greatly expand the molecular content of the currently used 17-marker miR*Inform*® Luminex-based assay. The Informagen workflow includes new formats for DNA and RNA quantification, an improved procedure for high-density primer multiplexing, and a multi-pool RT-PCR and tagging PCR step that enables simultaneous interrogation of DNA and RNA biomarkers of thyroid tumorigenesis and progression—all assessed in a single MiSeq sequencing run

CONCLUSIONS

The diagnostic value of the mRNA expression markers of the Informagen NGS Thyroid Test was demonstrated in 123 FFPE and 65 FNA samples. Importantly, the additional markers show improved diagnostic yield over the first-generation Luminex-based test. The Informagen NGS Thyroid Test illustrates the core components required to build a multi-categorical NGS assay that can be applied to other disease indications:

- The test combines well-established DNA and RNA markers with recently discovered and proprietary single nucleotide variations (SNVs), fusion transcripts, and mRNA expression markers integrated into a single assay.
- RNA fusion transcripts were detected with high sensitivity and specificity in FNA and FFPE samples and were analytically concordant with matched Luminex and whole transcriptome RNA-Seq data.
- Analytical concordance with matched whole transcriptome RNA-Seq data for mRNA expression markers in FFPE samples was achieved.

The results of this study demonstrate the feasibility and clinical value of combining DNA variants, RNA fusions and mRNA expression markers in a single test.

INTRODUCTION

Thyroid cancer is the most common endocrine cancer, and its incidence is increasing faster than any other cancer. Our understanding of the genetic events involved in thyroid cancer pathogenesis has allowed for the identification of a number of diagnostic, potentially prognostic and therapeutic targets. Several studies, including a study performed at Asuragen (Beaudenon et al.), have shown that molecular testing of fine-needle aspirate (FNA) biopsies for *BRAF, RAS, RET/PTC* and *PAX8/PPARγ* significantly improves the accuracy of the preoperative FNA diagnosis by cytology. Testing for these markers is formally recommended by the American Thyroid Association (ATA) to help resolve indeterminate or suspicious cytology and reduce unnecessary surgeries. These markers were the basis of Asuragen's first-generation thyroid test, miR*Inform*® Thyroid, a Luminex-based liquid bead array method for detecting DNA variants in *BRAF, HRAS, NRAS* and *KRAS* along with fusion transcripts from RET/PTC1, RET/PTC3 and *PAX8/PPARγ*. In 2010, Asuragen began offering clinical testing of thyroid FNA biopsies using miR*Inform*® Thyroid, and over 3000 samples have been tested with this platform.

Despite the clinical utility of these markers, studies have revealed that up to 30% of both papillary thyroid carcinomas (PTCs) and follicular thyroid carcinomas (FTC) do not carry any of the mutations or translocations in the aforementioned genes, leaving patients with an indeterminate diagnosis. Additional markers are needed to complement the information provided by the most common mutations/translocations and achieve the sensitivity and specificity necessary to significantly improve thyroid cancer diagnosis from FNAs. Furthermore, some markers have lower specificity as markers for malignancy, such as RAS and to a lesser extent, PAX8/PPARγ translocations, because they can also be associated with follicular adenoma (FA), a benign neoplasm that is considered a precursor to FTC. To improve the diagnostic performance of miR*Inform*® Thyroid, Asuragen identified additional biomarkers associated with cancer-positive, mutation-negative specimens (Table 1), leading to the development of a new thyroid cancer assay, the Informagen NGS Thyroid Test. This test will provide substantially improved diagnostic information, as well as information that will help prognosis and treatment decision-making.

Content	Analyte	Thyroid Test	
		miR*Inform*® (Luminex)	Informagen (targeted NGS)
Gene Panel	DNA	**4 genes** 13 regions *BRAF, KRAS, HRAS, NRAS*	**20 genes** 68 mutation-hotspot regions Additional content in *BRAF, KRAS, HRAS and NRAS*
Translocation Panel	RNA	RET/PTC1 RET/PTC3 PAX8/PPARγ	**52 gene fusions** 14 novel fusions 8 splice variants • 4 conferring resistance • 4 constitutively activating
Gene Expression Panel	mRNA	None	**40 mRNA** Iodine uptake markers • Predicts RAI efficacy • Markers of aggressiveness • Markers of progression Recurrent 3' fusion partners • Confirm fusions Thyroid marker control genes

Table 1. Expanded content of the multi-analyte Informagen NGS Thyroid Test. Assays are performed with NGS for DNA and RNA analytes. The Informagen content contains the ATA-recommended markers assessed by the miR*Inform*® Test on the Luminex platform. (RAI-radioiodine therapy.)

STUDY DESIGN

Assay Overview

The Informagen NGS Thyroid Test enables an in-depth molecular characterization of thyroid cancer FNA and FFPE specimens through NGS analysis of hotspot regions across 20 thyroid cancer-involved genes, 52 RNA fusions and aberrant isoforms, and 40 mRNA disease-relevant expression markers (Table 1). The selection of DNA and RNA markers was performed through an extensive review of thyroid literature, mining of databases such as COSMIC and TCGA, and internal biomarker discovery efforts using RNA-Seq to uncover novel fusions and diagnostic expression markers (For a detailed description of the discovery effort and assay migration process, see Haynes et al., 2015, in preparation). The expression signature was discovered by applying RNA-Seq to FFPE tumor biopsies to discover novel gene expression biomarkers and fusion transcripts associated with malignancy. All mRNA

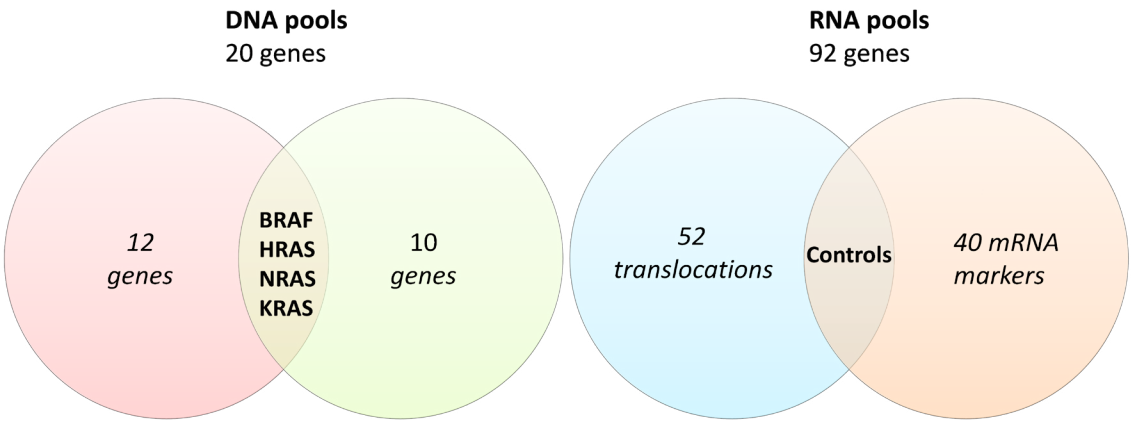

Figure 1. Summary of the molecular markers covered by the Infomagen test. Targets were detected using a 4-pool multiplex PCR-enrichment design, where each primer pool targets a specific class of marker. All amplified regions were sequenced simultaneously on a MiSeq NGS instrument. The assay was designed to cover 104 unique targets: 68 mutation-hotspot regions in 20 genes by DNA-Seq, and 52 translocations and aberrant isoforms and 40 mRNA expression markers by RNA-Seq.

marker amplicons were designed to span exon-exon junctions to enable compatibility with a residual gDNA background in a total nucleic acid (TNA) isolation. The signature was migrated to a diagnostic-ready platform, the Quantidex™ NGS Assay (Haynes et al., 2015), and was integrated with DNA markers under a single assay. An analytical and clinical performance assessment of the Informagen NGS Thyroid Test was performed on RNA and DNA control samples, as well as 164 FFPE clinical samples and 69 FNA clinical samples.

Assay Design

The DNA and RNA markers included in the Informagen thyroid panel are covered by four independent primer pools (Figure 1), the products of which are combined into a single pool for sequencing.

Pools 1 and 2: DNA-Seq targets. Targets were selected based on a comprehensive review of somatic mutations in COSMIC and TCGA for thyroid carcinomas. Targeted regions were ranked based on mutation prevalence in COSMIC and TCGA in thyroid carcinomas. 68 total regions were targeted, representing 20 genes and >1500 unique COSMIC mutations (v60 mutations represented across 2 primer pools). Mutational hotspots that are well-recognized as molecular diagnostic markers, including *BRAF* codons 600 and 601 and *HRAS*, *NRAS* and *KRAS* codons 12, 13 and 61, were covered by both Pool 1 and Pool 2 designs to enable even greater detection fidelity and built-in parallel confirmation to resolve these events.

Pool 3: Gene fusion targets. A total of 39 candidate fusion events in 24 independent biological specimens were identified in the whole transcriptome (WT) RNA-Seq FFPE cohort. 22 of the fusion-positive 24 specimens were malignant by histopathology. Primers were designed to independently confirm 35 of the 39 candidate fusions by endpoint PCR and Sanger sequencing. Sanger sequencing confirmed the presence of 23 of the 35 breakpoints, resulting in 18 verified unique fusion transcripts.

Pool 4: mRNA expression markers. mRNA expression markers covered by Pool 4 were informed by the WT RNA-Seq FFPE cohort on the basis of diagnostic potential and involvement in fusions previously known or identified in this study. Markers with theranostic or prognostic value were also included based on evidence from other studies. Similar to the DNA pools, Pools 3 and 4 contained overlapping targets for built-in assay fidelity and parallel confirmation.

Controls: Cell lines, Synthetic Targets, IVT RNA

Specific mixtures of the cell lines, synthetic variant and translocation targets, and in vitro transcribed products were used as controls for the development of the Informagen test. Cell lines with known variants were obtained from ATCC or other commercial entities. The DNA or RNA from these cell lines was isolated and molecularly characterized using targeted NGS analysis (Hadd et al., 2013). Purified DNA from the individual cell lines was combined to formulate pooled cell line DNA controls. A summary of available cell lines and mutation status is detailed in Haynes et al., 2015. In vitro transcription (IVT) was performed using a T7 high-yield transcription kit (MEGAshortscript). The RNA product from the IVT reaction was then mixed at the expected copy number

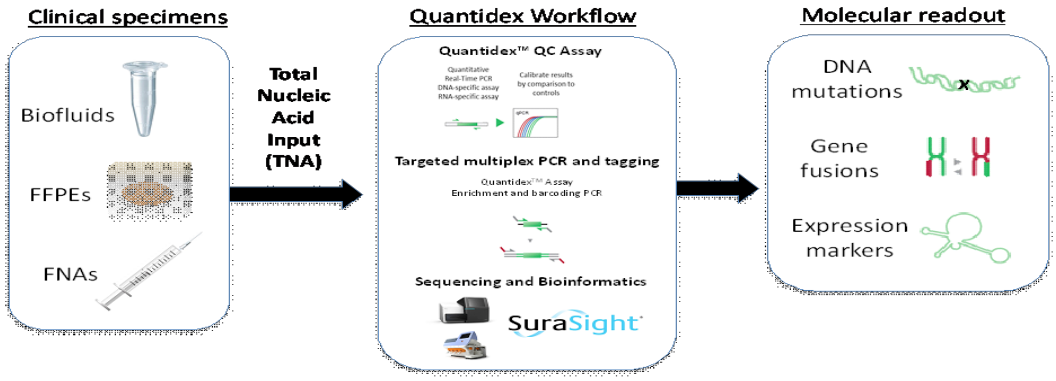

Figure 2. The Quantidex™ NGS-based assay provides a comprehensive solution for quantification of a unified RNA and DNA assay. The Quantidex™ NGS workflow integrates Quantidex™ sample QC for library prep and informatics on the most challenging clinical samples.

ratio into a background of wild type thyroid RNA.

Clinical Specimens

FFPE blocks of surgically resected thyroid lesions (n=69) and associated clinicopathological information were obtained from Asterand (Detroit, MI). For each specimen, a hematoxylin- and eosin-stained slide was prepared and reviewed by an independent pathologist at Asuragen to confirm the histologic diagnosis. Five 10-micron sections were then cut for nucleic acid isolation. Total nucleic acids (TNA) were extracted using the Ambion RecoverAll Total Nucleic Acid Isolation Kit for FFPE tissues (Life Technologies, Carlsbad, CA) according to the recommended protocol. The amount of DNA and RNA present in the derived TNA material was determined using the Quantidex™ DNA Assay to quantify the number of functional DNA copies and the Quantidex™ RNA Assay to quantify the number of RNA copies. The derived TNA and RNA material was then assayed for known thyroid mutations including *BRAF* V600E, *HRAS, NRAS, KRAS*, PAX8/PPARγ, *RET/PTC1* and *RET/PTC3* using the Luminex-based miR*Inform*® Thyroid Test (Table 1).

Thyroid FNA biopsies were processed according to internal protocols for high-throughput TNA extraction from FNA or small-input tissue stored in RNARetain.

Workflow

The Quantidex™ NGS workflow supports sequencing on Illumina's MiSeq system for the detection of DNA variants, RNA fusion transcripts and mRNA expression markers. A simplified schematic of the workflow, from sample acquisition to data analysis, is shown in Figure 2 and has been described in other publications (Hadd et al. 2013, Sah et al. 2013, Haynes et al. 2015). The data summarized in this paper describe multiple experiments performed with analytically defined control samples and well-characterized, residual clinical FFPE and FNA samples. Positive variant call thresholds for FNA samples were 2% variant and SuraScore >500, while the threshold was set at 5% variant and SuraScore >500 for FFPE samples. A detailed description of these data analysis protocols is provided in Haynes et al., 2015.

RESULTS

Analytical Assessment of the Informagen Test

Consistent with our previous experience in applying Quantidex™ QC Assays to evaluate the suitability of FFPE RNA and DNA for successful amplification, we found the Quantidex™ DNA and RNA QC Assays to be predictive of post-sequencing data quality in FFPE samples.

For the FFPE cohort, DNA Pools 1 and 2 showed 100% sensitivity with respect to mutations covered by the Luminex assay (data not shown). All mutations in *BRAF* and *RAS* called positive by Luminex were also called positive by the Informagen test. Concordant calls were made for 74/79 samples (93.7%, Figure 3). All five discrepancies were positive calls made by NGS and negative calls made by Luminex. The five positive Informagen calls were called positive in both Pools 1 and 2, including a 5% *BRAF V600E* variant, which is a clinically defined thyroid cancer marker. Interestingly, 7 *RAS* mutations were identified by Informagen (by both Pools 1 and 2) but were not called by Luminex. Considering the percent variant at which these *RAS* mutations were detected (>10%) and that all 7 specimens were malignant, it is highly probable that most of these *RAS* mutations are true positives. This finding further suggests a sensitivity advantage of the Informagen test over the Luminex assay in the context of poor-quality FFPE samples.

For the 3 fusions covered by both the Luminex assay and Pool 3, we observed 100% sensitivity and 100% specificity. Moreover, 13 fusions (excluding transplicing events) that were identified by whole transcriptome RNA-Seq were correctly called positive by the Informagen test, including several newly discovered fusions.

The Pool 4 expression data for the FFPE discovery cohort were compared to matching whole-transcriptome RNA-Seq data.

figure 4 table

Figure 4. Concordance between platforms. The scatter plots show concordance between NGS of RNA biomarkers with whole-transcriptome analysis.

Most mRNA markers demonstrated concordance between the two platforms. Agreement between the two platforms was notably worse for markers that were lower in expression (Figure 4).

For the 69-sample FNA clinical cohort, the Informagen test demonstrated concordance with Luminex calls for mutations covered by both platforms, and a number of novel mutations and fusions were identified predominantly in the malignant specimen. The Informagen test showed 100% sensitivity and 100% specificity for fusions covered by the Luminex assay.

Pool 4 mRNA expression markers were evaluated by PCA biplot analysis using the FNA cohort (n=50) and a merged FFPE cohort (n=X) (Haynes et al., 2015). Benign specimens were separated from malignant specimens for both FFPEs and FNAs, and FFPEs could be distinguished from FNA samples.

The resulting performance of the Pool 4–derived classifiers in both FFPE and FNA samples suggests strong diagnostic potential for the Pool 4 mRNA signature. Strikingly, the classifier that was trained on the FFPE-merged cohort and tested on FNAs produced results almost identical to the classifier that was trained on FNAs and tested on FNAs under cross-validation, highlighting the robustness of the classifier. Finally, the subset of 52 FNAs that were indeterminate by cytology was examined in isolation under the FNA-trained classifier and the FFPE-trained classifier (Figure 5). The observed performance in the cytologically indeterminate subset of FNAs both reaffirms the performance of the Pool 4 classifier and confirms its potential clinical utility in this subset of thyroid biopsies most relevant to molecular diagnostics.

Clinical Assessment of Informagen FNA Cohorts

Results from 50 cytologically indeterminate FNAs were assessed using both the Luminex miR*Inform*® panel (17 markers, Table 1) and the Informagen test. Thirteen (26%) samples were negative by the miR*Inform*® Thyroid Test, resulting in an indeterminate diagnosis. When these 50 samples were tested using the comprehensive thyroid NGS panel (Table 1, Figure 2), all but 1 of the 13 negatives could be diagnostically reconciled Figure 6. Of the 12 that could be

Classifier/ Training Cohort	Test Cohort	AUC	Sens	Spec	PPV	NPV	Accuracy
Pool 4 (0.5 threshold)/ FNA	FNA	0.87	0.78	0.79	0.83	0.73	0.78
Pool 4 + Luminex mutations (calibrated threshold) / FNA	FNA	0.91	0.95	0.64	0.78	0.9	0.82

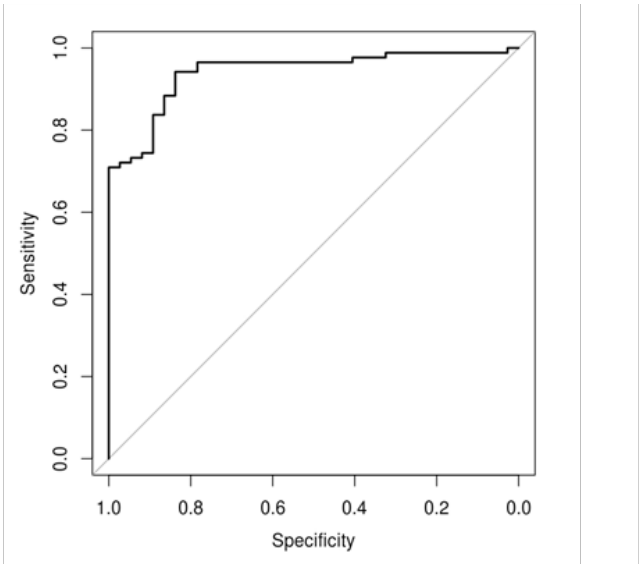

Figure 5. Summary of Pool 4 diagnostic classifier results over the FNA cohorts. Two thresholds were examined: a naïve threshold of 0.5 and a threshold that was calibrated under leave-one-out cross-validation with a target sensitivity of 95%.

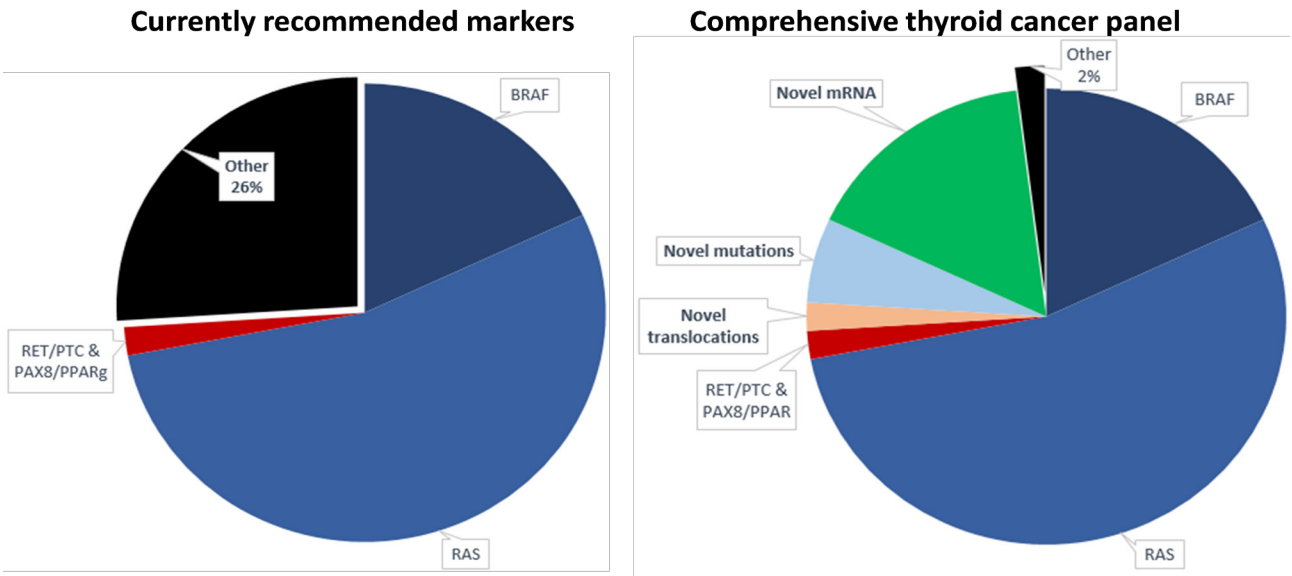

Figure 6. A unified DNA and RNA thyroid cancer panel can identify malignancies in cytologically indeterminate specimens that are undetected by conventional markers. The Informagen test demonstrated improved capability for the detection of thyroid cancer over National Comprehensive Cancer Network (NCCN)-recommended markers alone.

reconciled, 8 were by mRNA, 3 by mutation, and 1 by novel translocation, highlighting the clinical value of using a multi-categorical approach.

CONCLUSION

The purpose of this study was to develop the most advanced and comprehensive assay for the molecular diagnosis of thyroid cancer by combining multi-categorical testing of DNA variants, fusion transcripts and mRNA markers into a single NGS-based workflow. This goal was accomplished through the design, development, analytical verification and clinical accuracy of the Informagen assay, which profiles over 1500 known somatic DNA mutations, 52 gene fusions and 40 mRNA expression markers of diagnostic and theranostic value.

Compared to the first-generation Luminex-based test, the Informagen test demonstrated expanded coverage of known driver mutations, adding value in identifying driver mutations in malignant samples where no mutation was previously characterized. Similarly, the assay identified rearrangements in FNA and FFPE samples that were negative by the Luminex assay. Finally, the mRNA expression markers that were discovered by whole transcriptome RNA-Seq and pursued through targeted RNA-Seq showed potential to correctly classify the majority of the cytologically indeterminate mutation-negative FNAs with high specificity.

These results confirm the analytical performance of the Informagen test and reveal strong potential for its use as a diagnostic assay. Ultimately, the use of targeted NGS systems in validated clinical laboratories will be required to translate the outcomes of clinical trials into new therapies that improve health economics and patient outcomes.

Note: Interpace Diagnostics™, LLC, a subsidiary of PDI, Inc., acquired the miR*Inform*® and Informagen Thyroid Tests from Asuragen.

REFERENCES

1. Beaudenon-Huibregtse S., et al., *Thyroid*, 2014. **24**: p. 1479-87.
2. Hadd A.G., et al., *J Mol Diagn,* 2013.**15**: p. 234-47.
3. Haynes B., et al. 2015. [in preparation]
4. Sah S., et al., *Genome Med*, 2013. **5**: p. 77.

ASURAGEN, INC. 2150 Woodward Street, Suite 100 • Austin TX 78744 • tel 512 681 5200 • fax 512 681 5201 • asuragen.com