### Patent Analytics for Empowering Business Decisions

WIPO Regional Workshop on Patent Analytics WIPO-IPO of the Philippines Manila, Philippines December 4, 2013

> Cynthia Barcelon Yang Director

Scientific Information & Patent Analysis Group Information & Analytics Sciences



Together we can prevail.

# **Bristol-Myers Squibb at a glance**

<u>Mission:</u> To discover, develop and deliver innovative medicines that help patients prevail over serious diseases.

World-class science with global reach and experience

- 28,000 employees in >90 countries
- \$17.6 B Net Sales in 2012
- ~ 8,000 people in R&D worldwide (10 major sites)
- \$3.9 B R&D investments in 2012

125 year (1887-2012) History of Innovation

- A leader in biopharmaceuticals
- Benchmark BioPharma Company





Sales by Therapeutic Areas





\*Forxiga is not approved in the U.S.

# **Global Manufacturing**



# **String of Pearls Strategy**



# The Importance of Intellectual Property (IP)



### Vital Role of Patent Analysis in Corporate R&D \*

Enable innovation & strategic business decision-making by providing value-add analysis of the patent literature to support:

**Scientific Research & Development** 

- State-of-the-art patent landscape
- **IP Procurement & Protection** 
  - Patentability determinations
  - Freedom-to-operate assessments
  - Validity opinions



**Business Development & Strategic Transactions** 

- Competitive analyses
- Due diligence String of Pearls Strategy

<u>\* Pharmaceutical Patent Analyst</u>, March 2012, Vol. 1, No. 1, pp 5-7 <u>http://www.future-science.com/doi/full/10.4155/ppa.12.1</u>



# **Increasing Trend in Patent Filings\***

#### Figure A.2.1.1: Trend in PCT applications for the top five origins







Note: 2012 data are WIPO estimates.

Source: WIPO Statistics Database, March 2013



Keeping up with the flood of data and information is challenging.

Spotting patterns and extracting useful and actionable information is even harder.

# Patent Data Challenges

"how to "capture, explore and capitalize"

**"70-90% of information contained within patents is never published anywhere else".** 

- U.S. Office Tech Assessment & Forecast Report

Bristol-Myers Squibb

# **Emerging Technologies Opportunities**



#### Set of tools to facilitate knowledge extraction from patents

- Semantic 'concept-based' vs keyword-based
- Automated extraction/indexing & creation of virtual compounds from Markush structure claims in chemical patents

#### Tools for integrating multiple sources of data/information

 Patent alerts from patenting authorities (eg, USPTO PAIR, EP Register) & commercial sources (eg, CAPLUS, Inpadoc)

> See also <u>Text Mining & Visualization Tools – Impressions of Emerging Capabilities</u>, Yang et.al., World Patent Information, Vol. 30, (2008), pp 280-293.



### **Case Studies**

#### **Company Profiling\* to identify:**

- Technology areas
- Key researchers
- Patenting trends
- Potential for String-of-Pearl strategy implementation

#### **Technology Assessment to identify:**

- Industry trends in kinase assay technology platforms
- Competitive landscape of pharma organizations
- Trends in therapeutic areas and kinase family groups
- Business investment strategy direction & implementation

\* <u>Enhancing Patent Landscape Analysis with Visualization Output</u>, Yang et. al., World Patent Information, Vol. 32 (2010), pp. 203-220



### Case Study 1:Company Profiling of Kosan Biosciences

#### Main Objective:

 Assess Kosan Biosciences patent assets and research activities for potential String of Pearls engagement

#### **Questions:**

- Who are their top inventors? What are their research teams composed of?
- What is their research focus in terms of Mechanism of Action?
- What is their research focus in terms of Utility?



### Case Study 1: Company Profiling of Kosan Biosciences

Patent Analytics Tool:

VantagePoint



Sources:

Type of Search:

Patents Retrieved :

Derwent World Patent Index Chemical Abstracts Services

**Patent Assignee** 

**123 patents** 



#### Company Profiling Case Study: Kosan Biosciences Who are Kosan's top inventors? What are their research teams composed of?





#### What research has Kosan focused on in terms of Mechanism of Action ?





# What research has Kosan focused on in terms of <u>Utility</u>?





# **Research Landscape**





#### **Publication Year Trend**

**Key Researchers** 



### Case Study 2: Competitive Kinase Assay Technology Platform Analysis

#### Main Objectives:

- Assess competitive landscape of kinase assay technology platforms for drug screening to guide future investment & strategy at BMS
- Identify current assay technology trends used for different kinase groups/families in various therapeutic areas

#### **Questions:**

- What are the kinase assay technology trends?
- What assay technology platforms are being used by companies?
- What are the trends for kinase groups/families?
- What are the trends for different therapeutic areas?
- What are the trends in therapeutic area vs. kinase groups/families?

### **Case Study 2: Kinase Technology Platform Analysis**

Patent Analytics Tool:

Sources:

**Type of Search:** 

Patents Retrieved :

Linguamatics I2E

PatBase (Bibliographic data) IBM computer-curated patents (Full-text WO, EP, US patents)

**Key Concepts** 

> 7000 patents



# **Patent Analytics Approach – Overall Process**



### **Step 1 : Data Collection & Optimization**





### **Step 1: Data Collection & Optimization Clients Requirements & Search Strategies**

#### Collect client requirements

- Data coverage 2004 to 2011
- Trends reviewed in 2-year blocks 4 two-year blocks for this analysis

#### Generate relevant patent dataset

- PatBase Search kinase AND inhibitor terms near each other
- Limit to families with a US or EP or WO member
- Identify kinase assay technology terminologies
  - Search broad assay terms in Description section only
  - Get clients involved
  - Refine terms to get the most relevant retrieval

### Data Collection & Optimization Generating Patent Family Sets

#	Search query (edited)	Results
<u>1</u>	Title, Abstract, Claims=(kinas* w10 (inhibit* OR activat* OR modulat* OR antagon*))	12983
<u>2</u>	1 AND Country Code=(us OR ep OR wo)	11466
<u>3</u>	2 AND Description=(assay* OR bioassay* OR screen* OR measur* OR detect*)	10751
<u>4</u>	Patent families in which 1 <sup>st</sup> publication is in 2004 or 2005	1678
<u>5</u>	Patent families in which 1 <sup>st</sup> publication is in 2006 or 2007	1822
<u>6</u>	Patent families in which 1 <sup>st</sup> publication is in 2008 or 2009	1972
7	Patent families in which 1 <sup>st</sup> publication is in 2010 or 2011	1597

**Bristol-Myers Squibb** 

### Data Collection & Optimization Getting Full Text XML from IBM to I2E

### •Over 7000 PNs obtained from PatBase

- US (79%)
- PCT (20%)
- Other (1%)

### •Export 1 PN per patent family

Using PatBase family table format

#### •Retrieve patent full text xml from IBM internal Database

- PatBase xml quality
- Getting full text xml from IBM internal database
- "Patent Handler" In-house web-based utility that automatically pulls full-text xml from IBM internal database and index them into I2E server







### **Step 2 : I2E Query Development**





### Linguamatics I2E - Interactive Information Extraction



#### **Linguamatics I2E Query Development**

Indexed Patents on I2E server 1672 patent documents for 2004 – 2005 1839 patent documents for 2006 – 2007 2002 patent documents for 2008 – 2009 1597 patent documents for 2010 – 2011 >7000 patents total

**Query Development** Designed and driven by our desired I2E Output

#### **Designed Output Columns (partial) in**

Kinase Group	Kinase Group Kinase		Technology Clusters	Technology SYN			
			L				

#### Kinase group macro

- 500 Kinases with over 10 K synonyms in 10 groups
- Kinase groups for trends analysis

#### **Technology cluster macro**

- Terms provided by clients
- Multiple iterations to get the best possible results

#### Therapeutic area macro

- 5 therapeutic areas of interests
- I2E Disease Ontology

#### Patent assignee (major pharma) macro

STN company thesaurus



### **I2E Query Development - Technology Macro**





### Screenshot of I2E Query: Kinase Technology Terms





#### Screenshot of I2E Query: Kinases by Therapeutic Area





One therapeutic area =10 single queries - one per kinase group 5 therapeutic areas of interest

= 50 queries for all therapeutic areas and all kinase groups which are then combined



### Step 3: Analysis & Visualization of Results





### I2E Results: Table of "Assertions"







### **I2E Results: Export to Excel**

PN	PY	PP	PA	Title	Abstract	KG	Kinase	Kinase	TA	Disease	Disease	Doc	Hit
								SYN		Class			
US2011030113	3 2011	2010		COMPOSITION	The present invention relates to novel	XYZ	TRKA	TrkA	CNS	Demyelinatin	demyelin ation	1	injury relating to dysmyelination or
		2011			compounds and pharmaceutical					g Diseases	auon		disorder or condition associated
					compositions containing the same that are								with irregular TrkA activity in a patient
				KINASE	canable of inhibiting or antagonizing a family								with fregular the activity in a patient,
				INHIBITORS	of recentor tyrosine kinases. Tronomysosin								
					Related Kinases (Trk) in particular the nerve								
					growth factor (NGF) receptor. TrkA								
					3								
US2011030113	3 2011	2010		COMPOSITION	The present invention relates to novel	XYZ	TRKA	TrkA	CNS	Pain	pain	1	and/or prevention of pain, cancer,
		2011		S OF PROTEIN	synthetic substituted heterocyclic								restenosis, or condition associated
				RECEPTOR	compounds and pharmaceutical								with irregular TrkA activity in a patient,
				TYROSINE	compositions containing the same that are								
				KINASE	capable of inhibiting or antagonizing a family								
				INHIBITORS	of receptor tyrosine kinases, Tropomysosin								
					Related Kinases (Trk), in particular the nerve								
					growth factor (NGF) receptor, TrkA								
1100044000470	0.0044	0040			<b>T</b> I (1) (1) (1) (1) (1)	1000	510	ED!/	0110		10.1		
052011029473	2011	2010			I ne present invention provides a method and	XXX	Erk2	EKK	CNS	Multiple	a multiple		of an autoimmune disease such a
		2011			composition for the treatment and prevention					<u>Scierosis</u>	scierosis		multiple scierosis which is mediated
	<u> </u>			FOR THE	or an autoimmune disease such a multiple								by autoreactive of administering at







### **Analysis and Visualization - Deliverables**

#### What are the kinase assay technology trends?



#### **Fluorescence Activity**

- Predominant technology and growing

#### Radioactive

- Being replaced (old?)

#### Caliper

- Not changing much

#### ADP

- Just starting to grow (new?)

Note: Percentages on Y-axis are calculated from Spotfire data





# **Analysis and Visualization - Deliverables**

#### What kinase assay technology platforms are being used by companies?



### **Analysis and Visualization - Deliverables**

#### What are the trends for kinase groups/families ? What are the trends for different therapeutic areas?

#### **Therapeutic Area**

Oncology is the major therapeutic area for kinase use and is increasing
CV and Metabolics are decreasing
Immunology and CNS did not significantly increase over time



•No clear trend in kinase families

•Kinase Group 10 is the most important group and this is consistent with its role in cellular proliferation that is critical for Oncology and Immunology indications





### Analysis and Visualization - Deliverables Kinase Group Trends in Immunology

#### What are the trends in kinase groups/families for each therapeutic area?



# Case Study 2 – Data Summary

- > 7000 Full Text Patents; estimate half million pages of full text
- 510 Kinases with >10000 kinase synonyms
- >110 technology terms (including trademarks)
- **5** Therapeutic areas (CV, CNS, Met, Oncology, and Immunology)
- **3** Custom-built macros
- **60** Single I2E queries and **2** I2E multiqueries
- **12000** Rows of data in Excel (after cleaning up)
- ~ 2 GB interactive data in Excel with HTML source data plus Spotfire visualization packages

# **Business Value & Impact**

#### Business Value

- Manual if possible estimate 1 hour/patent
- 7000 hours or 875 days or 3.5 years for a FTE
- I2E Efficiency gain: 90%



#### •Business Impact – Client Feedback

- Analyzed results were relevant to the questions asked.
- Analysis provided key competitive assay technology platforms
- Allowed implementation of investment strategy 6 months ahead of schedule



# **Success Factors**

#### **Collaboration with clients**

 Multiple iteration in refining searching strategy, technology terms



#### i2e Tool - Strengths

- Powerful term extraction with ontology in various regions (fields)
- Use of macro to extract client defined data
- Highly interactive Excel output for easy analysis and drilldown
- Massive data extraction with multi-queries

#### Use of Spotfire for data visualization

Repetitive description for similar information was removed

### **Outcomes**

Provided key actionable insights that empowered business decisions.

Encouraged closer collaborations with R&D teams

 Recognized with BMS 3-I Award (Innovate, Integrate, Improve) Sept. 26, 2013

"Natural Language Processing to Mine Unstructured Text to Support R&D"





# **Patent Analyst's Role**

Select appropriate database sources and types of analysis and visualization tools that are most appropriate to the query and dataset based on:

- Scientific expertise
- Business knowledge
- Understanding of clients' needs
- Knowledge of tools and databases

Collaborate interactively with users to refine the query & analysis criteria and output

Guide users in navigating the dynamic reports to realize the full value of the report



### Achieving Patent Analytics Service Excellence



**Collaborative & Continuous Learning Culture** 

# **Acknowledgements**

- BMS Patent Analysis Group Information & Analytics Sciences (IAS) Research IT & Automation (RITA) Lead Evaluation & Mechanistic Biochemistry (LEMB)
- Linguamatics I2E
- Search Technology, Inc. VantagePoint
- **Chemical Abstracts Service STN Anavist**
- **Tibco Software, Inc. Spotfire**
- Minesoft, Inc. PatBase
- **IBM Text Analytics Consortium**



### Thank you!

### **Maraming Salamat Po!**

#### **Gracias!**

### Xie Xie!

Contact: Cynthia Barcelon Yang Director, Scientific Information & Patent Analysis Group Bristol-Myers Squibb Princeton, New Jersey USA Phone: 609-818-5515 Email: cynthia.yang@bms.com

