Beyond scholarly literature; a lens into innovation

Osmat A. Jefferson
QUT Library, May 8, 2014
Acknowledgments

- Arun Sharma (leadership at QUT)
- The Lens team and Richard Jefferson
  - With Support from Bill & Melinda Gates Foundation, Gordon and Betty Moore Foundation, QUT, The Lemelson Foundation, NCBI/NIH, USPTO, Qualcomm, and NICTA

- Deniz Köllhofer
- Thomas Ehrich
- Small Multiples

- Ben Warren, Nina Prasolova, Prabha Ajjikuttira, and Lance De Vine
- Research assistants and volunteers

- Ignacio Medina Castello  [www.genomemaps.org](http://www.genomemaps.org)
- Patent offices who responded to our questionnaire.
Patent-based innovation

Discovery  Invention

Patent

innovation

Patent Bargain

Use in the market

???

Contribution to Society
In his dissenting statement, Justice Mayer clarified:

“… he[applicant] must go beyond simply informing the public of the bare outlines of the invention. He must also tell what he believes to be the best embodiment of the invention, and he must do so in a way that allows the public to practice that embodiment.

This prevents the inventor from obtaining patent protection with a minimal disclosure that reveals only inferior forms of the invention, while retaining for himself the most advantageous modes of carrying the invention into practice.”

Glaxo Inc. v. Novopharm Ltd., 52 F.3d 1043, 1055 (Fed. Cir. 1995)
Invention # Innovation

- **1900**: One inventor → One invention → One innovation
- **1980**: inventor +1 → A few inventions → One innovation
- **2006**: entity +1 → 500s inventions → One innovation
  - entity +1 → 100s inventions
Evolution of the patent system

Thousands


Total US Patent Applications
Total US Patent Grants

(http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm)
Globally

• Expansion of patent holdings in emerging economies
• TRIPS 1995 and other treaties
• Novel business practices on the use of patents
• Strong movement towards collaboration among patent offices and harmonization of patent procedures around the world.
• Lack of transparency
• Uncertainty for innovators
Use of IP

“Today innovation is increasingly driven by collaboration and integration. If we do not evolve our IP use, IP may become extinct... if we do not evolve IP use and step away from the exclusion, eventually the entire IP system may lose its societal trust and be done.”

Michael Kock, Head of Intellectual Property, Syngenta on January 17, 2013
Frederick Sanger, Two-Time Nobel-Winning Scientist, Dies at 95

BY DENISE GELLENE

Dr. Sanger, a British biochemist, showed how amino acids link together to form insulin in 1958, and invented a method of "reading" molecular letters that make up the genetic code in 1980.

Or, copy and paste this URL into your browser: http://nyti.ms/17pkneG

November 20, 2013
Structure of a patent document

- Title/cover page
- Review of literature (prior art)
- Objects of the Invention
- A summary of the Invention
- Detailed information
- “Invention”
- Working Examples
- Claims
- Drawings or illustrations
- Sequence Listings
SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 124

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 11385 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
   (A) ORGANISM: Homo sapiens

(ix) FEATURE:
   (A) NAME/KEY: CDS
   (B) LOCATION: 229..10482

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GOTGGGCGGA GCTGTCTGAA CTAGGCAGCA GAGGCGAGC CGGCTGCGCA CTGCTGCGCC 60
TCTGCTGCGC TCGGGGTTTC TTTTGGCGCG GTGCTGCGCC GCGGGGAGAA GCGTGAGGGG 120
ACAGATTTTG GACCGCGCGG GCTTTTGCTCA GCTTACTCCG GCCAAAAAG AACTGACCT 180
CTGGAGCGGA CTTAGTTACC AAGCATTGGA GGAATATCGT AGGTAAAAA ATG CCT ATT 237
   Met Pro Ile 1

GGA TCC AAA GAG AGG CCA ACA TTT TTT GAA ATT TTT AAG ACA CCG TGC 285
Gly Ser Lys Glu Arg Pro Thr Phe Phe Glu Ile Phe Lys Thr Arg Cys
   5 10 15

AAC AAA GCA GAT TTA GGA CCA ATA AGT CCT AAT TGG TTT GAA GAA CTT 333
Asn Lys Ala Asp Leu Gly Pro Ile Ser Leu Asn Trp Phe Glu Glu Leu
   20 25 30 35

Cambia

enabling innovation

7/05/2014

Lens.org
ALGORITHM 1
SCORE 4687.86 bits (12158)
IDENTITIES 2313 / 2325 (99.5%)
POSITIVES 2318 / 2325 (99.7%)

<table>
<thead>
<tr>
<th>Query</th>
<th>Hit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2232</td>
<td>1951</td>
</tr>
<tr>
<td>KAFMEDD</td>
<td>KAFMEDD</td>
</tr>
<tr>
<td>DELT</td>
<td>DELT</td>
</tr>
<tr>
<td>DSKLP</td>
<td>DSKLP</td>
</tr>
<tr>
<td>SHATH</td>
<td>SHATH</td>
</tr>
<tr>
<td>SLFTCP</td>
<td>SLFTCP</td>
</tr>
<tr>
<td>NEEE</td>
<td>NEEE</td>
</tr>
<tr>
<td>MVLNSRIGK</td>
<td>MVLNSRIGK</td>
</tr>
<tr>
<td>RRGPEPLILVG</td>
<td>RRGPEPLILVG</td>
</tr>
<tr>
<td>2281</td>
<td>2000</td>
</tr>
<tr>
<td>1951</td>
<td>2000</td>
</tr>
<tr>
<td>2282</td>
<td>2282</td>
</tr>
<tr>
<td>EPSIKRNLLN</td>
<td>EPSIKRNLLN</td>
</tr>
<tr>
<td>EFDR</td>
<td>EFDR</td>
</tr>
<tr>
<td>IEQ</td>
<td>IEQ</td>
</tr>
<tr>
<td>KSLKASKSTP</td>
<td>KSLKASKSTP</td>
</tr>
<tr>
<td>DGI</td>
<td>DGI</td>
</tr>
<tr>
<td>TKIDRRLF</td>
<td>TKIDRRLF</td>
</tr>
<tr>
<td>MHHSVSEPI</td>
<td>MHHSVSEPI</td>
</tr>
<tr>
<td>2331</td>
<td>2331</td>
</tr>
<tr>
<td>2051</td>
<td>2051</td>
</tr>
<tr>
<td>2381</td>
<td>2381</td>
</tr>
<tr>
<td>CYPFRTT</td>
<td>CYPFRTT</td>
</tr>
<tr>
<td>KER</td>
<td>KER</td>
</tr>
<tr>
<td>QIEIQPNFTA</td>
<td>QIEIQPNFTA</td>
</tr>
<tr>
<td>PGQF</td>
<td>PGQF</td>
</tr>
<tr>
<td>LSKSH</td>
<td>LSKSH</td>
</tr>
<tr>
<td>LYE</td>
<td>LYE</td>
</tr>
<tr>
<td>HL</td>
<td>HL</td>
</tr>
<tr>
<td>LEKS</td>
<td>LEKS</td>
</tr>
<tr>
<td>SSLNLAVS</td>
<td>SSLNLAVS</td>
</tr>
<tr>
<td>GHP</td>
<td>GHP</td>
</tr>
<tr>
<td>2100</td>
<td>2381</td>
</tr>
<tr>
<td>2051</td>
<td>2051</td>
</tr>
<tr>
<td>2381</td>
<td>2381</td>
</tr>
<tr>
<td>CYPFRTT</td>
<td>CYPFRTT</td>
</tr>
<tr>
<td>QIEIQPNFTA</td>
<td>QIEIQPNFTA</td>
</tr>
<tr>
<td>PGQF</td>
<td>PGQF</td>
</tr>
<tr>
<td>LSKSH</td>
<td>LSKSH</td>
</tr>
<tr>
<td>LYE</td>
<td>LYE</td>
</tr>
<tr>
<td>HL</td>
<td>HL</td>
</tr>
<tr>
<td>LEKS</td>
<td>LEKS</td>
</tr>
<tr>
<td>SSLNLAVS</td>
<td>SSLNLAVS</td>
</tr>
<tr>
<td>GHP</td>
<td>GHP</td>
</tr>
<tr>
<td>2100</td>
<td>2100</td>
</tr>
<tr>
<td>2381</td>
<td>2381</td>
</tr>
<tr>
<td>2381</td>
<td>2381</td>
</tr>
<tr>
<td>CYPFRTT</td>
<td>CYPFRTT</td>
</tr>
<tr>
<td>QIEIQPNFTA</td>
<td>QIEIQPNFTA</td>
</tr>
<tr>
<td>PGQF</td>
<td>PGQF</td>
</tr>
<tr>
<td>LSKSH</td>
<td>LSKSH</td>
</tr>
<tr>
<td>LYE</td>
<td>LYE</td>
</tr>
<tr>
<td>HL</td>
<td>HL</td>
</tr>
<tr>
<td>LEKS</td>
<td>LEKS</td>
</tr>
<tr>
<td>SSLNLAVS</td>
<td>SSLNLAVS</td>
</tr>
<tr>
<td>GHP</td>
<td>GHP</td>
</tr>
<tr>
<td>2100</td>
<td>2100</td>
</tr>
<tr>
<td>2381</td>
<td>2381</td>
</tr>
<tr>
<td>2381</td>
<td>2381</td>
</tr>
</tbody>
</table>

E VALUE: 0E0
LENGTH: 2325 aa
Query: S 99.7%
Hit: C 68%
- **GenBank/NCBI**: 11,203,821 (72,147 docs)
- **USPTO/Fulltext-GreenBook**: 450,524 (18,696 docs)
- **USPTO/Fulltext-RedBook**: 7,997,628 (163,133 docs)
- **USPTO/Bulk Listings**: 119,461,910 (6,826 docs)
RAW TEXT

Raw data processing
- Sequence listings extracted
- Format checked and harmonised
- Each sequence fingerprinted
- Metadata collected, if available (organism, sequence length, type, etc)
- Spellchecking of organism names and assignment of taxonomic id

Redundancy analysis based on
- Jurisdiction
- Data source
- Applications/grants
- Sequence length

Data merging multiple data sources per sequence listing are being tracked.

DBase Storage

Faceted analysis index

BLAST Sequence search

Taxonomy-based sequence search

Sequence search

Homology based clustering

Homology Screening

NCBI RefSeq

Patent Search Index

 SEQ ID NO

Full-text locations

Patent Data

PatSeq Data Inventory

PatSeq Data UI

Non Redundant View

Genome Mappings

BWA / BLAT

Alignment Datastore

PatSeq Explorer

PatSeq Analyzer

PatSeq Data UI

PatSeq Data UI

PatSeq Data UI

PatSeq Data UI