Basic principles of IG sequence analysis: 
**Immunogenetic analysis: in vitro**

Lesley Ann Sutton  
Dept. of IGP, Uppsala University, Sweden  
Dept. of Molecular Medicine & Surgery, Karolinska Institutet, Stockholm, Sweden  
Lesley.sutton@igp.uu.se
Different nature of IG genes: 
**IG genes = a unique set of genes**

**Non-IG genes**

![Diagram showing TP53 gene structure with polymorphisms c.639A>G p.R213R; VAF=50% and c.337T>G p.F113V; VAF=80%]

**Polymorphism c.639A>G p.R213R; VAF=50%**

**c.337T>G p.F113V; VAF=80%**

**Structure:**

Single gene – one → many exons

**Variations:**

- SNP
- Pathogenic mutations

**Origin of variants:**

- Inherited
- Acquired
Different nature of IG genes:

**IG genes = a unique set of genes**

![Diagram of B cell receptor (BcR) with labeled regions: V region, C region, Antigen binding site, H chain, L chain, IG, CD79a/b, Signalling subunits.](image-url)
Different nature of IG genes:

**IG genes = a unique set of genes**

**IG genes**

**IGH genes**

**IGL genes**

**Structure**
- Rearranged IG gene
- Loss of nucleotide
- Non-templated nucleotides

**Variations**
- SNP

**Generation of diversity**
- Acquired
Generation of diversity:

**IG genes = a unique set of genes**

1. **Combinatorial diversity**
   - V region encoded by 2 or 3 genes
   - Reservoir of multiple IG V, D, J genes
   - Random assembly

2. **Junctional diversity**
   - Imprecise joining at the CDR3

3. **Combinatorial diversity**
   - Pairing of heavy and light chain IG genes

4. **Maturation diversity**
   - Somatic hypermutations (SHM)

---

Central
(bone marrow)

Peripheral
(2° lymphoid organs)
Germline organization of the IGH locus

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>V subgroups</th>
<th>D</th>
<th>J</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>55</td>
<td>7</td>
<td>23</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>ORF</td>
<td>7</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>46</td>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>108</td>
<td>7</td>
<td>27</td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>
**IG V(D)J recombination**

*IG genes = a unique set of genes*
**IG V(D)J recombination**

*IG genes = a unique set of genes*
**IGH V(D)J recombination**

**IG genes = a unique set of genes**
**IG V(D)J recombination**

**IG genes = a unique set of genes**
IG V(D)J recombination

IG genes = a unique set of genes
IGHV  IGHD  IGHJ  IGHM

5'  germline DNA
     D-J rearrangement

     partially rearranged DNA
     V-D rearrangement

     rearranged DNA

     precursor mRNA

     mature mRNA

     polypeptide chain

     Elimination of the signal peptide

     Mu heavy chain

Transcription

Splicing

Translation
Junctional diversity: created by processing of the coding ends

Modification due to:
- Nucleotide trimming
- Palindromic nucleotides
- nontemplated nucleotides

- RAG cleavage
- Hairpin opening
- Deletion
- Exonuclease activity
- $N$ nucleotide addition
- Terminal deoxynucleotidyl transferase (TdT)
- Ligation

Modification due to:
- Nucleotide trimming
- Palindromic nucleotides
- nontemplated nucleotides
From the bone marrow to the periphery
Antigen exposure further shapes the repertoire

Kuppers, Nat Rev Cancer, 2005
Diversification in the mature B cell repertoire

Several modes of modification

1. **Somatic hypermutation (SHM)**
   - V region
   - Mostly single base substitutions
   - Exceptionally high mutational frequency: $10^{-3}$ to $10^{-4}$/base/cell division

2. **Class switch recombination (CSR)**
   - C region
   - Localized: start near the 5’ end of the V-D-J-EXON and extend for 1-2 kb
   - Preferred target hotspot motifs
   - Transitions more frequent than transversions

3. **Switch from mIGs to secreted Igs**
   - Silent (S) or replacement (R) or stop codon
Diversification in the mature B cell repertoire

**Several modes of modification**

- **Insertions/Deletions**
  - Rare events (3%)  Subset #2 (25%)
  - Productive? If reading frame is maintained
    - Mostly single base substitutions
    - Exceptionally high mutational frequency: \(10^{-3}\) to \(10^{-4}/\text{base/cell division}\)
    - Localized: start near the 5’ end of the V-D-J-EXON and extend for 1-2 kb
    - Preferred target hotspot motifs
    - Transitions more frequent than transversions
    - Silent (S) or replacement (R) or stop codon

Considerations for protocol optimization

IGHV mutational analysis - *optimization at two levels*

- technical protocols - generate a *reliable IGH sequence*
  - avoid missing sequences
  - avoid incorrect sequence data

- clinical interpretation - generate a *reliable report*
  - correct interpretation/implications
From the patient to an IG sequence:

important parameters

Material:
- Cell source
- Anticoagulant
- Work-up of cells

Type of nucleic acid

PCR methodology
- PCR protocol
- *Taq* polymerase
- PCR primers

Processing & clonality

Sequencing
IG gene analysis in CLL

Material: *Cell source*

- peripheral blood
- bone marrow
- lymph node
- other

**IGHV SHM status**

- stable feature irrespective of the leukemic cell source
- stable throughout the disease course

Survey from previous workshop (n=63)
IG gene analysis in CLL

Material: \textit{Anticoagulant}

- EDTA tubes (Ethylene diamine tetra-acetic acid)
- CPT tubes (Citrate/pyridoxal 5’-phosphate/tris)
- heparinized tubes
- other

Survey from previous workshop (n=63)
IG gene analysis in CLL

Material: **Work-up of cells**

- Ficoll gradient (PB / BM)
- cell suspension of biopsy (e.g. for flow analysis)

Survey from previous workshop (n=63)

Other:
- FFPE (2), whole blood (4), red cell lysis (2), osmotic analysis (1)
IG gene analysis in CLL

**Type of Nucleic acid**

- genomic DNA (gDNA)
- RNA / complementary DNA (cDNA)

Survey from previous workshop (n=63)

Quantity of gDNA / RNA
- mostly 100-500 ng gDNA (range 1 ng – 1 ug)
- mostly 1 ug RNA for cDNA reaction (range 400 ng – 2 ug)
### IG gene analysis in CLL:

**Type of Nucleic acid**

<table>
<thead>
<tr>
<th>Molecule</th>
<th>advantages</th>
<th>disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>gDNA</td>
<td>- more optimal for long-distance transport</td>
<td>- non-productive rearrangement can</td>
</tr>
<tr>
<td></td>
<td>- use of archival material</td>
<td>- also be amplified</td>
</tr>
<tr>
<td>RNA/cDNA</td>
<td>- identifies mostly only productive rearrangement</td>
<td>- reverse transcription step required</td>
</tr>
<tr>
<td></td>
<td>- allows isotype identification</td>
<td></td>
</tr>
</tbody>
</table>

→ no scientific rationale for choosing gDNA or RNA/cDNA

→ advisable to use similar type of nucleic acid in multi-center trials
IG gene analysis in CLL:

**Type of Nucleic acid**

Unproductive rearrangements are not restricted to gDNA!

9.1% cases carried unproductive rearrangements

<table>
<thead>
<tr>
<th>Type of Nucleic acid</th>
<th>RNA/cDNA</th>
<th>gDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unproductive rearrangements</td>
<td>1.6%</td>
<td>13.9%</td>
</tr>
<tr>
<td>Single unproductive rearrangements</td>
<td>0.4%</td>
<td>0.8%</td>
</tr>
</tbody>
</table>

Langerak et al. Leukemia 2011
IG gene analysis in CLL:

**Type of Nucleic acid**

Double rearrangements are not restricted to gDNA!

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unproductive rearrangements</td>
<td>377/4154</td>
</tr>
<tr>
<td>cDNA</td>
<td>26/1628</td>
</tr>
<tr>
<td>gDNA</td>
<td>351/2526</td>
</tr>
<tr>
<td>Single unproductive rearrangements</td>
<td>27/4154</td>
</tr>
<tr>
<td>cDNA</td>
<td>7/1628</td>
</tr>
<tr>
<td>gDNA</td>
<td>20/2526</td>
</tr>
<tr>
<td>Double rearrangements</td>
<td>435/4154</td>
</tr>
<tr>
<td>cDNA</td>
<td>61/1628</td>
</tr>
<tr>
<td>gDNA</td>
<td>374/2526</td>
</tr>
<tr>
<td>Productive + unproductive</td>
<td>350/435</td>
</tr>
<tr>
<td>cDNA</td>
<td>19/350</td>
</tr>
<tr>
<td>gDNA</td>
<td>331/350</td>
</tr>
<tr>
<td>gDNA</td>
<td>325/350</td>
</tr>
<tr>
<td>Similar mutational status</td>
<td>18/326</td>
</tr>
<tr>
<td>cDNA</td>
<td>18/326</td>
</tr>
<tr>
<td>gDNA</td>
<td>17/350</td>
</tr>
<tr>
<td>Similar mutational status</td>
<td>17/350</td>
</tr>
<tr>
<td>cDNA</td>
<td>16/350</td>
</tr>
<tr>
<td>gDNA</td>
<td>17/350</td>
</tr>
<tr>
<td>Unmutated productive + mutated unproductive</td>
<td>308/326</td>
</tr>
<tr>
<td>cDNA</td>
<td>16/350</td>
</tr>
<tr>
<td>gDNA</td>
<td>17/350</td>
</tr>
<tr>
<td>Mutated productive + unmutated unproductive</td>
<td>7/350</td>
</tr>
<tr>
<td>cDNA</td>
<td>7/350</td>
</tr>
<tr>
<td>gDNA</td>
<td>7/350</td>
</tr>
<tr>
<td>Double productive</td>
<td>30/435</td>
</tr>
<tr>
<td>cDNA</td>
<td>13/85</td>
</tr>
<tr>
<td>gDNA</td>
<td>17/85</td>
</tr>
<tr>
<td>Similar mutational status</td>
<td>18/56</td>
</tr>
<tr>
<td>cDNA</td>
<td>18/56</td>
</tr>
<tr>
<td>gDNA</td>
<td>18/56</td>
</tr>
<tr>
<td>Discordant mutational status</td>
<td>28/85</td>
</tr>
<tr>
<td>cDNA</td>
<td>15/29</td>
</tr>
<tr>
<td>gDNA</td>
<td>15/29</td>
</tr>
</tbody>
</table>

### Abbreviations:
cDNA, complementary DNA; gDNA, genomic DNA.

RNA/cDNA: 3.8%

RNA/cDNA: ≈ 1/3

gDNA: 14.8%

gDNA: ≈ 2/3

Langerak et al. Leukemia 2011
IG gene analysis in CLL

PCR methodology: **PCR primers**

- **IGHV leader primers**
  - LH (family-specific) – CH (*Sahota, Blood 1996*)
  - LH & VH (family-specific) – CH (*Fais, J Clin Invest 1998*)

- **IGHV FR1 primers**
  - FR1 consensus – JH (*Aubin, Leukemia 1995*)
  - FR1 multiplex – JH (BIOMED-2) (*Van Dongen, Leukemia 2003*)

- **IGHV FR2 primers**: short IGHV

- **IGHV FR3 primers**: too short IGHV sequences

- **Downstream primers**: IGHJ or isotype specific IGHC
IG gene analysis in CLL

PCR methodology: **PCR primers**
IG gene analysis in CLL

PCR methodology: *PCR primers*

Primer sets

- **IGHV leader primers** $\rightarrow$ ERIC 2017 UPDATED RECOMMENDATIONS

- **IGHV FR1 primers** $\rightarrow$ *only in RARE CIRCUMSTANCES*
  - FR1 consensus – JH (*Aubin, Leukemia* 1995)
  - FR1 multiplex – JH (BIOMED-2) (*Van Dongen, Leukemia* 2003)

- **IGHV FR2 primers**: short IGHV sequences $\rightarrow$ *NOT ACCEPTABLE*
  (only recommended upon negative leader / FR1 results, due to SHM)

- **IGHV FR3 primers**: too short IGHV sequences $\rightarrow$ *NOT ACCEPTABLE*
IG gene analysis in CLL

**Clonality Testing**

Preferred methods for clonality testing

- heteroduplex analysis
- high resolution PAGE
- Gene Scan / fragment analysis
- agarose gel electrophoresis: too low resolution → **DISCOURAGED**

<table>
<thead>
<tr>
<th>Strategy</th>
<th>advantage</th>
<th>disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD analysis / PAGE</td>
<td>- unlabeled products</td>
<td>- lower detection limit</td>
</tr>
<tr>
<td></td>
<td>allows direct sequencing</td>
<td></td>
</tr>
<tr>
<td>GS analysis</td>
<td>- higher detection limit</td>
<td>- labeled products less optimal in sequencing</td>
</tr>
<tr>
<td></td>
<td>- optimal visualization</td>
<td></td>
</tr>
</tbody>
</table>
Heteroduplex analysis in clonality testing: 
*interpretation and sequencing strategy*

**What is the next step?**
- direct sequencing
- single PCR → sequencing
- gel excision → sequencing
- cloning → sequencing

**Patients A, C, E, F, G**
- monoallelic
  - direct sequencing

**Patient D**
- bi-allelic
Heteroduplex analysis in clonality testing: interpretation and sequencing strategy

Patients A, C, E, F, G

- monoallelic

Patient D

- bi-allelic

What is the next step?

- direct sequencing
- single PCR → sequencing
- gel excision → sequencing
- cloning → sequencing
Heteroduplex analysis in clonality testing: *interpretation and sequencing strategy*

Patient B: What are the possible scenarios?

a. Small clone below detection limit → follow-up sample
b. Clone is present, but not recognized → different primer set required
c. Lymphocytosis not due to CLL clone → check immunophenotype
Genescan analysis in clonality testing: interpretation and sequencing strategy

**Case 1:** monoallelic → direct sequencing in principle possible

**Case 2:** monoallelic + background → 1. direct sequencing ? ; 2. gel excision

**Case 3:** polyclonal → 1. no clone (no CLL ?); 2. small clone below detection (follow-up sample ?)
IG gene analysis in CLL: Sequencing strategies

Direct sequencing

• one product from multiplex PCR
  → starting from IGHJ / IGHC, then reverse sequence via IGHV family primer
• one product from single PCR
  → sequencing via specific primers from both sides

Sequencing after gel excision + elution

• bi-allelic rearrangement: physical separation of products

Sequencing after subcloning

• final option, e.g. physical separation of bi-allelic rearrangements impossible
**IG gene analysis in CLL**

**ERIC recommendations**

<table>
<thead>
<tr>
<th>Most important and relevant parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of nucleic acid</strong></td>
</tr>
<tr>
<td><strong>PCR primers</strong></td>
</tr>
<tr>
<td><strong>Clonality analysis</strong></td>
</tr>
<tr>
<td><strong>Sequencing</strong></td>
</tr>
</tbody>
</table>

→ Many parameters: no clear scientific rationale for one or the other option

→ Some strategies show complementary value

→ Clonality testing is an essential phase in the strategy!
**PROGNOSTICATION - IMMUNOGLOBULIN GENE ANALYSIS**

ERIC recommendations for IGHV gene mutational analysis in CLL (Shtal, Leukemia 2007)

**TP53 ABERRATIONS**

Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: Updated ERIC recommendations. (Rosenquist, Leukemia 2017)

ERIC recommendations on TP53 mutation analysis in Chronic Lymphocytic Leukemia (Hospislova, Leukemia 2012)
IGHV Gene Analysis Certification

**BACKGROUND**
CLL cells are endowed with a plethora of communication systems facilitating microenvironmental interactions.

**AIMS OF THE NETWORK**
ERIC aims to promote and/or advance the determination of IGHV gene mutational status in CLL for diagnostic and

**STRUCTURE OF THE NETWORK**
The IG Network consists of a total of 7 Reference Centres and 2 Certifying Centres.

**CERTIFIED CENTRES**
ERIC is proud to announce that it currently has 62 certified centres in over 25 countries.

**CERTIFICATION ROUNDS**
2 certification rounds have been completed. Certification rounds are held just twice a year.

**PARTICIPATION FORM**
Please complete the Participation Form which remains active for the whole year by clicking on read more.

http://www.ericll.org/ighv-gene-analysis-certification/
COLLECTION OF PROBLEMATIC CASES AND DATABASE CREATION

"Problematic cases" exist that defy a clear-cut classification and include those showing any of the following features:

- Double in-frame rearrangements
- Double in-frame rearrangements with discordant mutation status
- Unmutated in-frame rearrangement coexisting with a mutated out-of-frame rearrangement
- Unmutated heavy chain rearrangement associated with a mutated light chain rearrangement
- Single transcribed out-of-frame rearrangements
- Single transcribed rearrangements carrying stop codons
- Rearrangements carrying any kind of deletions or insertions/duplications
- In-frame rearrangements carrying a mutation either at IMGT codon C-104 (end of H\(\text{H}\)3) or at W-118 (start of H\(\text{H}\)4)
- Rearrangements without a clearly defined junction

Though these cases are probably known to anyone working in IGHV analysis in CLL, they are limited in frequency, hampering meaningful analysis at a single-institution level. Therefore, we propose to collect in a dedicated database all problematic cases in order to reach a significant number of them. Everyone in the CLL community is welcome to participate to this joint effort; to submit problematic IGHV sequences and participate to the project please go to the submission form.

COORDINATION

- Lesley Ann Sutton - lesley.sutton@guu.se
- Anastasia Hadzidimitriou - anastasia.xtz@gmail.com

USEFUL LINKS FOR IMMUNOGLOBULIN GENES ANALYSIS

http://www.ericll.org/tools/
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 692298.